TetR-binding aptamer as a versatile regulatory element

Mol, Adam Artur

(2020)

DOI (TUprints): https://doi.org/10.25534/tuprints-00011845

Publikationstyp: Dissertation

Fachbereich: 10 Fachbereich Biologie

Quelle des Originals: https://tuprints.ulb.tu-darmstadt.de/11845

TetR-binding aptamer as a versatile regulatory element

vom Fachbereich Biologie
der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Dissertation
von Adam Artur Mol

Erstgutachterin: Prof. Dr. Beatrix Süß
Zweitgutachterin: Prof. Dr. M. Cristina Cardoso

Darmstadt 2018

Mol, Adam Artur: TetR-binding aptamer as a versatile regulatory element
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2020
URN: urn:nbn:de:tuda-tuprints-118455
Tag der mündlichen Prüfung: 25.06.2018

heute ist ein schöner Tag

Acknowledgements

I WISH TO EXPRESS MY SINCERE GRATITUDE TO ALL WITHOUT WHOM THIS DISSERTATION WOULD NEVER COME TO EXIST

I was lucky to be a part of Süß research group and the MetaRNA ITN programme!

European Union's Horizon 2020 research and innovation MetaRNA programme under the Marie Skłodowska-Curie [642738]

I am extremely grateful to my PhD advisor **Professor Beatrix Süß**

- for her scientific advice, knowledge and many discussions and suggestions!
- for her kindness!
- for the freedom to clone ‘random’ constructs :) !

I am also very grateful to Professor Cristina Cardoso for helpful suggestions, access to FACS/microscopy facilities and friendly attitude!

Thanks to the members of the Reading Committee!

Haaalllooo!
Britta S. thank you for your huge help in performing experiments and great shared time in the lab! I wish you only ‘lazy days’ :D !

Thank you to all office mates! especially for Adri**a**n and Flo!
Thanks for making noise, your friendship and fun and ‘French air’ too!
Flo, I know nobody’s pissed you off like me :P ! oՈo

Cristina! Moltes gràcies for quick lunches and gossip and life advice to ser feliz cada dia! Los problemas no existen!

Anne & Julia W. thanks for your help with blotting and microscopic pics!

Anett, Britta K., Dunja, Michi, Julia, Janka … and all lab Team thanks for really great time!

Grazie Adolorada! :D

Jairo, Maryja, Monia, Sylwina, Zuz … and all my friends for support and friendship!
Agata thanks za herbatki, listening to my complaints and corrections!
Alex thanks for reading the boring thesis for you and your comments!
Sabine thank you for the opportunity to live in Reinheim (where this thesis also was written)!

and special thanks to my Mom, Dad, Lidii and all my family! Without the inspiration and support you have given me, I wouldn’t be the person I am today, I wouldn’t be where I am now!

- Thank you - Danke - Gracias - Dziękuję - Merci - Grazie - Obrigado –

Adam

Parts of this thesis are published in:

Mol, A. A., M. Vogel and B. Suess. 2020. **Inducible nuclear import by TetR aptamer-controlled 3' splice site selection.** RNA DOI: 10.1261/rna.077453.120

Mol, A. A., F. Groher, B. Schreiber, C. Rühmkorff and B. Suess. 2019. R**obust gene expression control in human cells with a novel universal TetR aptamer splicing module.** Nucleic Acids Research, 47: https://doi.org/10.1093/nar/gkz753

Contents

1 Summary

Synthetic biology explores the means of redesign and fabrication of existing biological systems or the *de novo* design and generation of biological components that are entirely new to nature. The main focus of the relatively new and dynamic discipline is to develop programmable genetic regulatory systems. Precise, reversible and temporary control of gene expression as well as in-depth understanding of fundamental genetics are crucial for the programming of new genetic circuits.

RNA represents one of the most powerful substrates in the engineering of biological systems, as it is versatile, designable and easily characterizable. Among the diverse functions of RNA molecules, their role as natural riboswitches has predominantly inspired researchers to design synthetic RNA-based regulators. Most of RNA devices contain a sensor element, an aptamer domain, which recognizes small molecules or protein ligands with high specificity and affinity, and an expression platform, controlling gene expression *via* various mechanisms. Generally, binding of a specific ligand to the aptamer domain stabilizes the RNA molecule or causes conformational changes in its structure; these further regulate transcription, translation and mRNA processing and degradation. Engineered RNA-based devices have already demonstrated multiple applications in synthetic biology. However, their implementation was mostly validated in bacteria and yeast, while mammalian synthetic biology has lagged behind.

Splicing of pre-mRNAs is an essential process in human cells that generates a diverse proteome through networks of coordinated splicing events and offers an additional layer of control for synthetic RNA devices. The reprogrammed removal of intronic sequences could provide a novel approach for the development of gene therapies to tackle disease phenotypes. For this purpose, it is necessary to design tools that allow precise and timely control of the splicing mechanisms. In the first research project described in the presented doctoral thesis, a versatile and highly efficient splicing device enabling control of gene expression in human cells and making use of an RNA aptamer recognized by the TetR was designed. Further, the portability of the splicing device was shown through its functionality in various reporter systems and the endogenous gene context. In the course of the thesis, the first inducible model for alternative 3’ splice site recognition with the tetracycline repressor (TetR) aptamer leading to production of splice variants with different subcellular localization was generated. The applicability of the system was corroborated in experiments aimed at controlling nuclear import in human cells. The proposed approach may prove valuable

in phenotypic studies of essential genes and provide an alternative in the development of therapeutic strategies, as the nucleo-cytoplasmic transport is vital for the maintenance of balanced cell physiology, with aberrant spatiotemporal localization of proteins leading to the development of various disorders and cancer. Finally, the third research project undertaken in the course of my doctoral studies focused on the development and optimization of the TetR aptamer dependent translational control system in human cells. Translational regulation constitutes an important point of post-transcriptional control of gene expression, enabling the cell to rapidly change the level of a specific gene product. Up to now, no efficient aptamer-ligand based translational regulatory system has been demonstrated in mammalian cells.

2 Introduction

2.1 Synthetic biology

Synthetic biology is a rising interdisciplinary research area, including biotechnology, genetic engineering, molecular biology, biophysics, electrical engineering, control engineering and evolutionary biology. One of the aims of the field is to engineer new living functionalities by creating, characterizing, and assembling biological parts, devices, and systems in living cells. The ability to re-engineer living organisms has immense potential to address societal needs with a number of applications, ranging from energy, to environment, and to medicine [1, 2].

Synthetic biology is based on scientific progress in biology which occurred over the past 50 years, especially DNA cloning, amplification and sequencing techniques as well as the transformation or transfection possibility to insert foreign DNA within a cell [3, 4] (Figure 2.1). The discovery that the rate of gene expression can be controlled by transcription factors, allows it to view genes as dynamical systems with inputs and outputs, where inputs and outputs are proteins. These parts can thus be assembled to form functional modules and larger systems. In fact, the properties of these components are often altered by the nucleic acid sequences of the genetic elements surrounding them. The design of the first two synthetic genetic circuits, a ring-oscillator and a toggle switch in the year 2000 was based on these technologies [5, 6]. At this time, much work was focused on the combination of a few DNA parts to form simple circuit modules with the aim of understanding the purpose of similar naturally occurring motifs. The synthetic biology has progressed to a “systems view” of biological processes, focusing on generating larger systems composed of well-characterized parts and subsystems. Theoretically, an ideal designed synthetic device should maintain its input/output behaviour as characterized in isolation unchanged upon combination with other modules and be transferable between bacteria to yeast and mammalian cells [7]. Failure of modularity leads to a long and repeated design process where subsystems are re-engineered from scratch any time a new module is added, thus presenting a challenging obstacle to scaling up circuits’ size. Currently, intense research has gone into strategies for enabling modular and layered design, what it is important to set the basis for the rational design of systems that are sophisticated enough to solve real-world problems. For this purpose, the scientists have made considerable efforts toward creating novel parts, characterizing existing parts,

providing insulation between modules, and enforcing functional circuit modularity against the effects of loads through the design of load drivers [1, 8, 9].

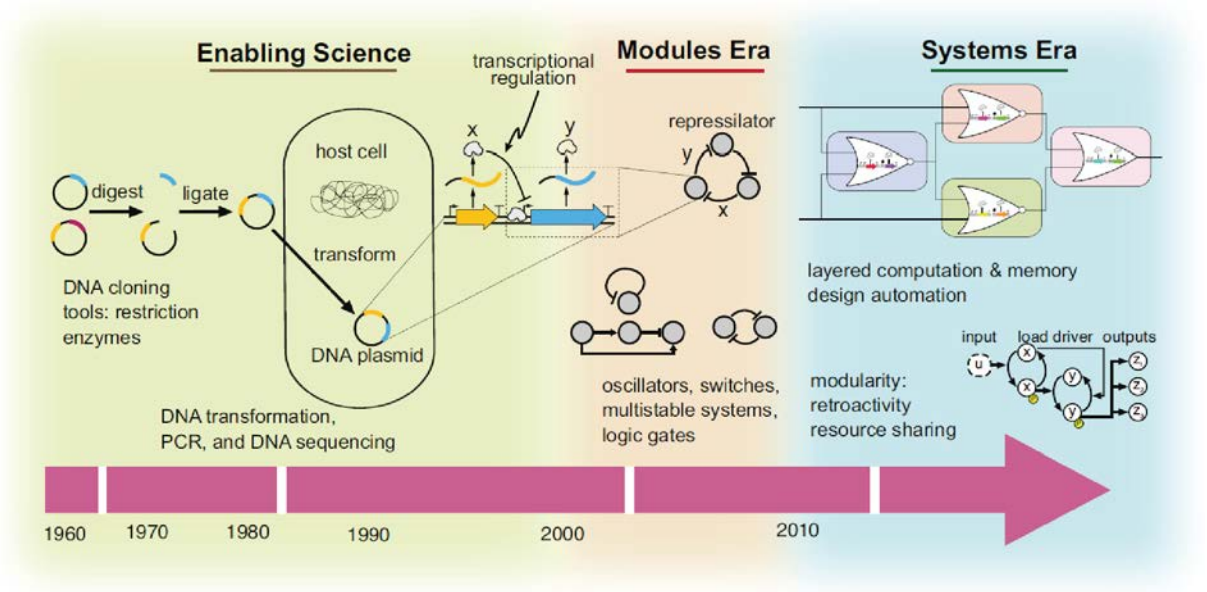


Figure 2.1 Synthetic biology on the temporal axis. (left) Enabling science: DNA cloning tools. (middle) Modules era: the first synthetic systems created were simple modules performing tasks such as oscillations and switching. (right) Systems era: construction of more complex circuits is based on a modular/layered design approach (adapted from [1]).

2.1.1 Applications of synthetic biology to health

Synthetic biology can revolutionize disease and cancer diagnosis and treatment. Synthetic devices can sense the intracellular concentrations of multiple molecular species, carry out logic computations through biomolecular reactions, and output a visible signal when a set of logic conditions are met [1, 10]. These logic conditions can be designed to classify the chemical signature typical of cancer cells, so that the circuit can recognize cancer and trigger a number of actions [11] (Figure 2.2A). Synthetic genetic circuits can program bacteria to colonize target sites where cancer is detected, providing a promising approach to reduce invasive tests for diagnosis and health monitoring [12, 13]. Programmed bacteria can further serve as smart vehicles for drug delivery by lysing at the tumour site and releasing therapeutic proteins to reduce tumour activity (Figure 2.2B). Synthetic biology also provides powerful tools to program immune T cells, to specifically attack cancer cells [14] (Figure 2.2C). Synthetic receptors engineered on T cells combined with biomolecular logic gates can identify cancer cells with high specificity.

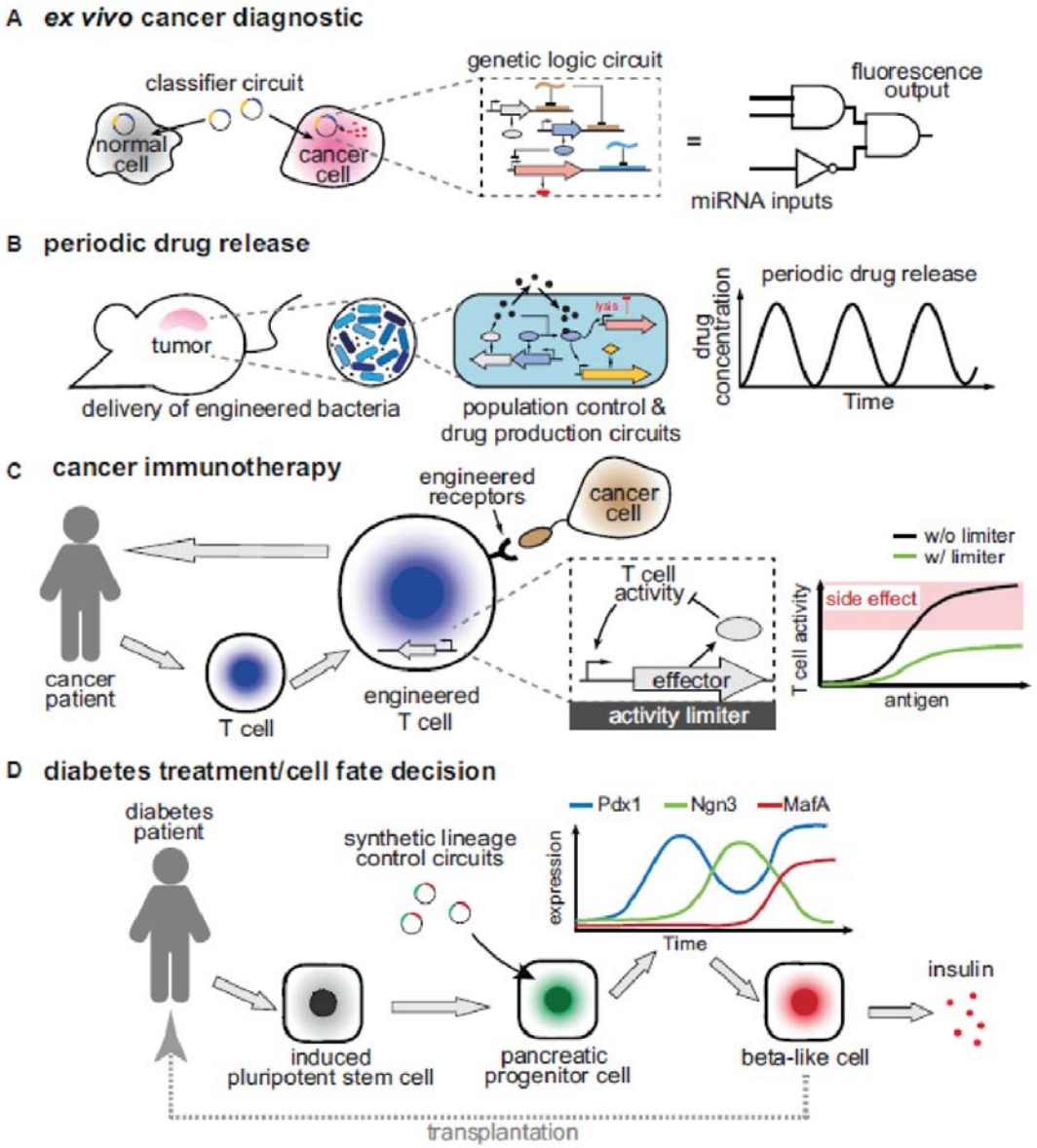


Figure 2.2 Applications of synthetic biology to health (A) A multi-input cell type classifier circuit used for cancer diagnostic *ex vivo*. A reference profile of miRNAs that are expressed in cancer cells is used to construct a genetic logic circuit realized through RNA interactions. When transfected into a cancer cell, the output of the logic circuit triggers expression of a fluorescence protein. **(B)** Bacteria can be engineered to periodically release a drug *in vivo*. A consortium of engineered bacteria is delivered to the target tumour site. Each cell contains a genetic clock, a cell lysis gene, a therapeutic protein production gene and a cell-cell communication module. The synchronized clocks control cell lysis in a periodic manner to release the therapeutic proteins, resulting in periodic drug delivery to the patient. **(C)** Synthetic genetic circuits increase the specificity and safety of cancer immunotherapy. Receptors can be engineered to trigger T cell activity when cancer cells are detected. Feedback loops can be used to limit the response of T cell activity to avoid side effects. **(D)** A synthetic lineage control circuit, activating the expression of three transcription factors according to a temporal pattern, hIPSCs can be reprogrammed into insulin-secreting beta-like cell for treating diabetes (adapted from [1]).

Synthetic controllers can then interact with the cellular chemotaxis pathway to migrate T cells to tumour cites. After T cells reach the target site, *in vivo* genetic controllers can actuate negative feedback actions to regulate the duration and strength of T cell activity to protect non-cancerous cells. Synthetic biology may also be used to enhance understanding of natural systems, including understanding of cell fate decisions and can provide tools to reprogram cell fate for regenerative medicine (Figure 2.2D). It was demonstrated that a synthetic reprogramming circuit that converts pancreatic progenitor cells derived from human induced pluripotent stem cells (hIPSCs) into insulin-secreting beta-like cells by strictly regulating the timing and expression of three key transcription factors *in vivo* [15]. To summarize, synthetic biology could prove remarkably effective in regenerative medicine where some damaged tissues and organs are traditionally replaced by biomaterials to restore proper function. These and many more tissues could instead be replaced by patient-derived cells that have been re-programmed through appropriate temporal and spatial control, avoiding innate immune responses.

2.2 The potential of aptamers as synthetic elements for regulating gene expression

2.2.1 *Trans*-regulatory RNAs

A main objective of synthetic biology is the specific control of cellular behaviour to understand fundamental genetics. Precise, reversible and temporary control of gene expression is necessary and can be achieved at the transcriptional, translation and post-translational levels [16, 17]. RNA is not only an intermediate for the conversion of gene expression to protein synthesis but also provides an attractive molecular scaffold for the design of genetic control elements. By interacting with ligands, RNA can alter its shape to act as a switch inside the cell [18, 19]. Moreover, RNA-based systems allow fast regulatory responses, genetic modularity and portability, features that make them suitable for complex platforms to achieve a broad spectrum of regulatory outputs [20, 21, 22]. RNA utilizes various mechanisms to control gene expression. One example of *trans* regulators represent non-coding RNAs (ncRNAs) which can have specialized functions in diverse biological processes and can shape nuclear structure and regulate gene expression. ncRNAs are divided into long (lncRNAs) and short (sRNAs) groupings. lncRNAs can act at various levels of gene expression and their functions are highly diverse. Many lncRNAs perform their functions by interacting with enhancers,

promoters, and chromatin-modifying complexes to regulate the transcriptional level. These interactions are mediated by specific short RNA sequence motifs or larger secondary or tertiary structures [23, 24]. Additionally, lncRNAs interacts with microRNAs (miRNA) networks in order to control gene expression at the post-transcriptional level. Small RNAs (sRNAs) that have been identified in bacteria, where they help to respond to changes in the environment [25, 26]. Generally, sRNAs regulate their target mRNAs through limited base pairing interactions, which leads to further changes in mRNA translation and stability, and consequently influences target gene expression. RNA-mediated regulations are also widespread in eukaryotes, where single-stranded miRNAs inhibit mRNA translation and double-stranded short interfering RNAs (siRNAs) can cleave mRNAs [27]. Another class of sRNAs represent Piwi-interacting (pi) RNAs. The piRNAs form complexes with Piwi proteins regulating transposable elements through an RNA-dependent DNA methylation mechanism in germline cells [28]. Moreover, the CRISPR-Cas9 system has provided another class of highly efficient tools to perform gene regulation with the use of guide RNAs [29].

2.2.2 *Cis*-regulatory RNAs

Riboswitches are an extraordinary example class, as they are *cis*-regulatory structured RNA elements capable of controlling expression of downstream genes by direct response to a small molecule ligand [30, 31]. The first riboswitch binding a flavin mononucleotide was discovered in the 5' UTR (untranslated region) of prokaryotic mRNA by Breaker [32]. Riboswitches include a ligand-sensing domain called an aptamer, and a regulatory domain, called the expression platform. Ligand binding to the aptamer stabilizes the aptamer structure and causes a conformational change or other activation mechanism in the expression platform, which mediates gene regulation [33]. In bacteria, riboswitches control transcription by forming a terminator or anti-terminator stem, or control translation by either occluding or exposing the ribosomal binding site (RBS or Shine-Dalgarno (SD) element) sequence that needs to be recognized by the ribosome during translation. A uncommon mechanism involves ligand-mediated ribozyme activation. It has been observed that ligand binding activates self-cleavage of the *glutamine-fructose-6-phosphate amidotransferase* RNA, leading to mRNA degradation [34].

So far it is still unclear how widespread riboswitches are. Riboswitch-based gene control is prevalent in bacteria, regulating approximately 2% of all genes in the bacteria

species *Bacillus subtillis* [35]. More than twenty distinct classes of natural aptamers are currently known that showcase the capability of the aptamers for selective and tight recognition of diverse compounds in cells. Many various genes are regulated by riboswitches. Most of riboswitches inhibit the production of unnecessary biosynthetic enzymes or transporters when a compound is already present at sufficient concentration. Moreover, some riboswitches can activate the expression of salvage or degradation pathways when their target molecules are present in excess [36, 37]. Although, most of these riboswitches have been identified in bacterial species, the coenzyme thiamine pyrophosphate-sensing riboswitches have been found in fungi and plants and it represents the most widespread riboswitch class discovered to date [19, 38, 39].

2.2.3 Riboswitches as genetic control devices

Engineered riboswitches known so far exploit RNA aptamers as a core component. Aptamers are capable of binding nearly any ligand of choice with high affinity and specificity and can be selected *in vitro* by directed evolution (SELEX: systematic evolution of ligands by exponential enrichment) which has led to the rapid and prevalent application of engineered riboswitches as artificial genetic control devices in synthetic biology over the past decade [18]. In classical SELEX, a pool of usually 10^{14} randomized sequences is mixed with an immobilized target. Non-binding molecules are removed by washing whereas bound molecules are eluted, amplified and subjected to next rounds of selection [33]. Gradually increasing the stringency during the following cycles may lead to aptamers that bind with high affinities and discriminate between closely related compounds. To date, aptamers targeting amino acids, proteins, small metal ion, organic molecules, bacteria, viruses, whole cells and animals have been generated. These aptamers have been widely applied in analytical, bioanalytical, imaging, diagnostic and therapeutic fields [40].

A major benefit of riboswitches is the possibility to combine sensing, transmitting and regulating domains within one molecule, enabling a direct approach to engineering synthetic devices with desired and defined functions. However, only a few aptamers have the properties that allow their exploitation as sensing domains of riboswitches [41 42]. Natural aptamer domains are conserved in secondary structure over bacterial phyla and the associated expression platforms can differ in both sequence and mode of

action between species. Therefore, engineering expression platforms requires choice of regulatory mode and fine-tuning of allosteric regulation, which often cannot be done rationally. The difficulty in design results from a limited information on ligand-free structures of riboswitches and also from a few full riboswitch structures that include the expression platform, as opposed to just the aptamer domain [19, 22].

In bacteria, engineered riboswitches most often target translation initiation, either by controlling access to the RBS through helix slippage or by sequestering the RBS [19] (Figure 2.3A). Regulation of gene expression by a theophylline-mediated translational switch efficiently allowed the control of chemotaxis in *Escherichia coli* [43]. However, only recently transcriptionally-based synthetic riboswitches have been documented and in this case a transcriptional terminator is formed depending on the binding state of the aptamer [44, 45, 46] (Figure 2.3B).

The cleavage activity of small ribozymes, like the hammerhead ribozyme have proved to be versatile tools in synthetic biology. Engineered allosteric aptazymes comprised of a self-cleaving ribozyme and an RNA aptamer have been exploited to achieve ligand-dependent regulation of gene expression via a variety of mechanisms in prokaryotic and eukaryotic cells [47, 48, 49]. In bacteria, the aptazymes were used to liberate the ribosomal binding site after ligand-dependent ribozyme cleavage (Figure 2.3C).

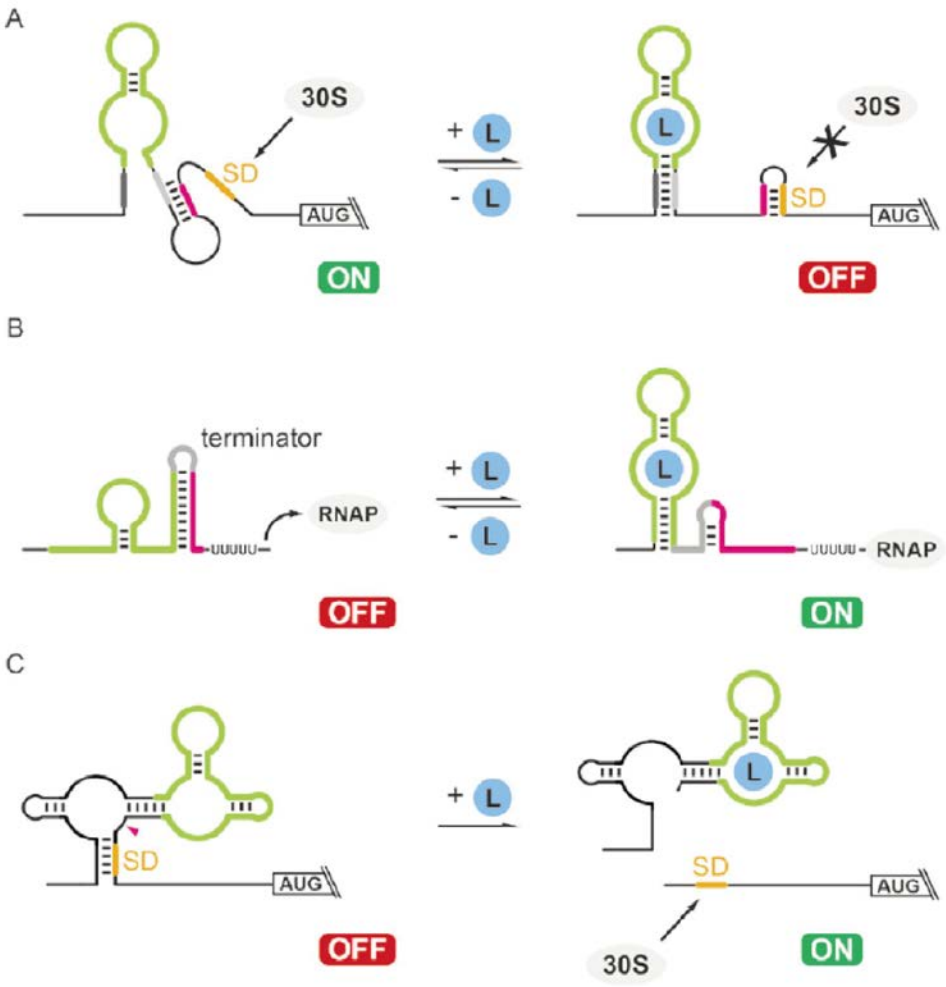


Figure 2.3 Mechanisms of engineered riboswitches in bacteria. (A) Regulation of translation. In the absence of a ligand, a stem-loop structure is formed between the aptamer domain (in green) and a sequence element complementary to the Shine-Dalgarno (SD) sequence (pink). Thus, the SD sequence (orange) is accessible for 30S binding and translation initiation occurs. As a consequence of ligand binding (in blue) and folding of the aptamer domain, an alternative stem-loop is formed which sequesters the SD sequence, blocking the binding of the 30S ribosomal subunit. **(B)** Regulation of transcription. The aptamer domain is fused to a short spacer region (gray), followed by a sequence complementary to the 3' part of the aptamer (pink) and a U stretch. In the absence of a ligand, the complementary 3' part is base-paired with the aptamer forming a terminator structure; thus, RNA polymerase (RNAP) dissociates and transcription is terminated. Upon ligand binding, terminator structure formation is inhibited and transcription can proceed, resulting in expression of the reporter gene. **(C)** Regulation with apatzyme. A ligand-dependent aptazyme is inserted into the 5' UTR of an mRNA in a way that ligand-induced self-cleavage liberates a sequestered SD sequence to induce translation initiation (adapted from [31]).

In eukaryotes, translational regulation by engineered riboswitches functions differently [33]. Insertion of the aptamer into a eukaryotic 5' UTR region leads to inhibition of translation initiation, either by preventing binding of the small ribosomal subunit to the mRNA cap structure or by interfering with ribosomal subunit scanning for the AUG start codon (Figure 2.4A). Several groups demonstrated translation control with aptamers which specifically recognize a Hoechst dye, malachite green, tetracycline, neomycin, biotin or theophylline [50, 51, 52, 53]. Additionally, a cap-independent translation initiation by an internal ribosome entry site has also been controlled by engineered riboswitches [54]. The only naturally occurring eukaryotic riboswitches have been discovered in filamentous fungi, green algae and higher plants where they control gene expression via regulation of pre-mRNA splicing [55, 56, 57, 39]. Analogously, theophylline- or tetracycline-binding aptamers have been used to block splicing by placing them close to either the 5' splice site, the 3' splice site or the branch point sequence [58, 59, 60, 61] (Figure 2.4B).

Aptazyme-based control of gene expression in yeast and mammalian cell lines using hammerhead ribozymes (Figure 2.4C) has been demonstrated by several groups [62, 63, 64]. Aptazyme cleavage at the 5' UTR leads to loss of the 5' cap, preventing ribosome initiation while aptazyme activity at the 3' UTR leads to loss of the poly-A tail, favouring degradation of the mRNA transcript. These aptazymes can activate or repress gene expression, depending upon whether the small molecule ligand turns the ribozyme on or off. Apart from the hammerhead ribozyme, the *hepatitis delta virus* ribozyme has also been successfully applied to RNA engineering [65].

RNA interference (RNAi) is an important and widely used mechanism in eukaryotes to repress gene expression. RNAi controls a broad range of developmental and physiological processes and it is also a standard system in molecular biology with a possible therapeutic strategy [66, 67]. Consequently, attempts to control the biogenesis pathway of RNAi were demonstrated with TetR or theophylline aptamers or aptazymes at sites of siRNA or shRNA molecules where their presence can interfere with processing by either Drosha or Dicer [68, 69, 70, 71] (Figure 2.4D).

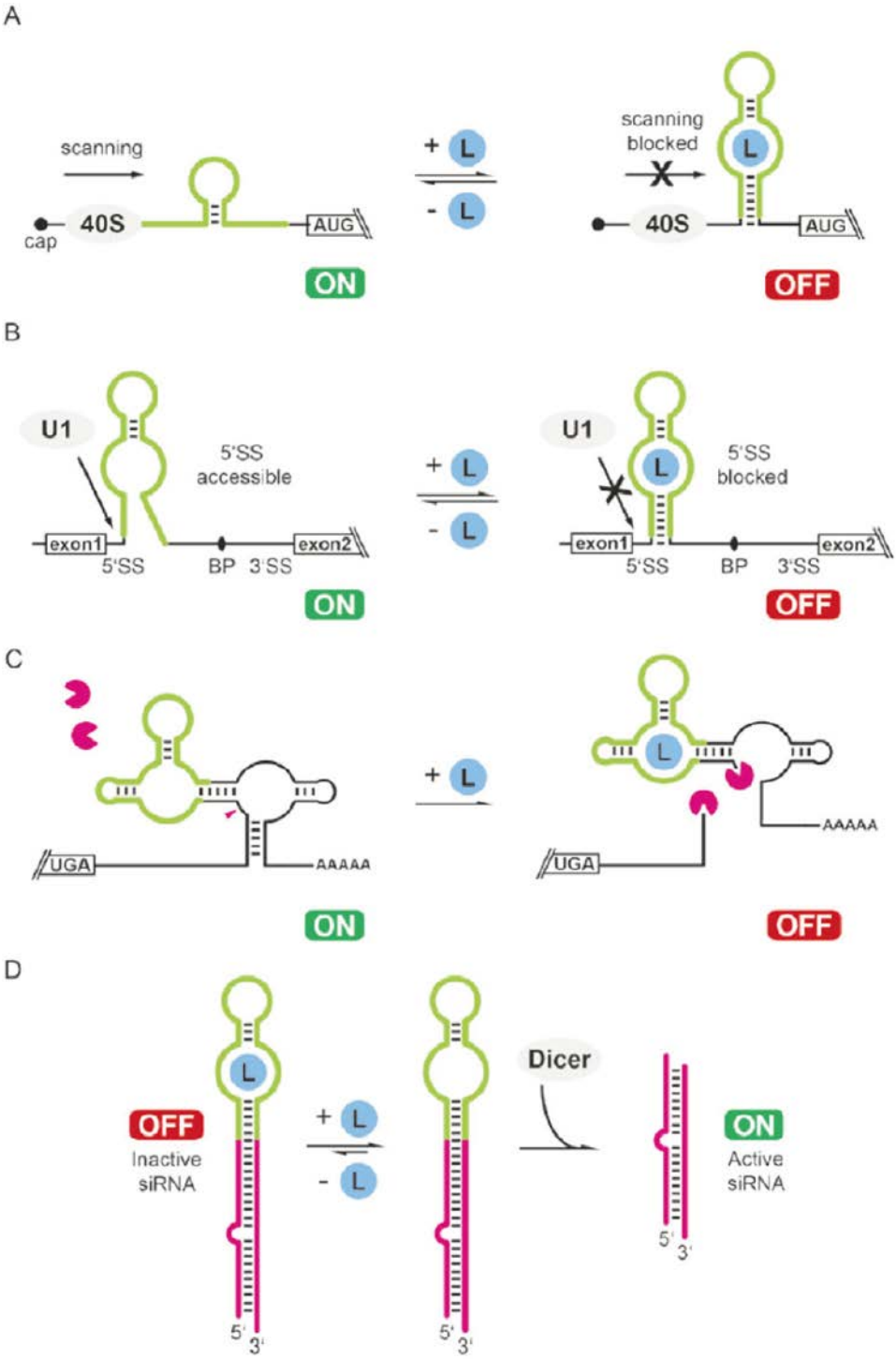


Figure 2.4 Mechanisms of engineered riboswitches in eukaryotes. (A) Regulation of translation. After insertion of the aptamer (in green) into the 5' UTR of a mRNA, ligand binding (in blue) to the aptamer prevents scanning of the 40S ribosomal subunit and, consequently, translation initiation. **(B)** Regulation of pre-mRNA splicing. An aptamer is placed into an intron of a eukaryotic mRNA to control the accessibility of the 5' splice site (SS). Ligand binding blocks splicing. **(C)** Regulation with aptazyme. A ligand-dependent aptazyme is inserted into the 5' UTR. Ligand-induced self-cleavage triggers RNA degradation (pink pacman). **(D)** Regulation of RNA interference. An integrated aptamer domain (green) interferes in its ligand-bound state with the enzymatic activity of Dicer (adapted from [31]).

2.2.4 RNA binding proteins (RBPs)-based synthetic RNA devices

Several successful attempts demonstrate conditional gene expression with synthetic RNA devices [31]. Among others, RNA binding proteins (RBPs) were used to predictably influence gene expression that both increases the number of cellular functions available for regulation while also enhancing the precision of the regulation [72]. One example of a protein-responsive RNA switch that controls translation in mammalian cells is based on the interaction between the ribosomal protein L7Ae and the box C/D kink-turn [72]. In other studies the L7Ae and the bacteriophage coat protein MS2 were used to design microRNA high- and low-sensors to engineer complex circuits in mammalian cells [73]. The next example of a RBP device was developed to control alternative splicing [74]. RNA structures recognised by proteins involved in NFkB and Wnt signalling were used to control cell fate through exon skipping. Thus, the expression of *herpes simplex* virus-thymidine kinase that confers sensitivity to a pro-apoptosis drug was controlled. Another interesting example of an RBP-based switch makes use of Pumilio and its derivatives, such as Pumilio and FBF (PUF) protein that can bind to RNA. The PUF domain has been fused with a splicing regulatory domain to regulate gene expression, primarily by controlling exon skipping [75], [76].

2.2.5 The tetracycline repressor (TetR) and the TetR-binding aptamer

The Tet systems in eukaryotes are derived from the transposon *Tn10* tetracycline (tc) resistance operon and it represents the most common method for inducing eukaryotic gene expression, for both, *in vivo* and *in vitro* applications. The prokaryotic Tet system was modified to be operative in eukaryotic cells [77]. It is based on the tc repressor (TetR), whose ligand is the antibiotic tc or several of its derivatives [68]. TetR is one of the most studied transcriptional regulators. It is genetically, biochemically and structurally well characterized and belongs to the Tet/Cam family of bacterial regulators. TetR controls the expression of the *Tn10* encoded tc resistance in Gram-negative bacteria. The actual resistance protein TetA is the an antiporter that actively export the tc out of the cell. The constitutive expression of TetA protein is toxic to the cell. Therefore, in the absence of this antibiotic TetR prevents the transcription of *tetA*. TetR is a 46.6 kDa homodimer with N-terminally located helix-turn-helix motifs [78], [79] (Figure 2.5A). In the absence of a ligand, TetR forms a dimer that strongly binds to the tet operator sequence (*tetO*). In the presence of a ligand TetR dissociates from the DNA, and gene expression is activated [80]. In mammalian cells, TetR can be used as

both a repressor (Tet-OFF), or converted into an activator (Tet-ON) by fusing it to the transactivation domain from virion protein 16 of the *Herpes simplex virus*. This system was initially developed for applications in model organisms where the tools for controlling gene expression are limited. More recently, homologues of TetR have been used to build synthetic gene switches for various applications and switches responding to other antibiotics [81]. The RNA aptamer recognized by the TetR was found by combining *in vitro* selection with *in vivo* screening in *Escherichia coli* [82]. The TetR aptamer folds into a stem-loop structure with an internal loop that displays the protein-binding site, while the P1 and P2 stems can be easily modified (Figure 2.5B).

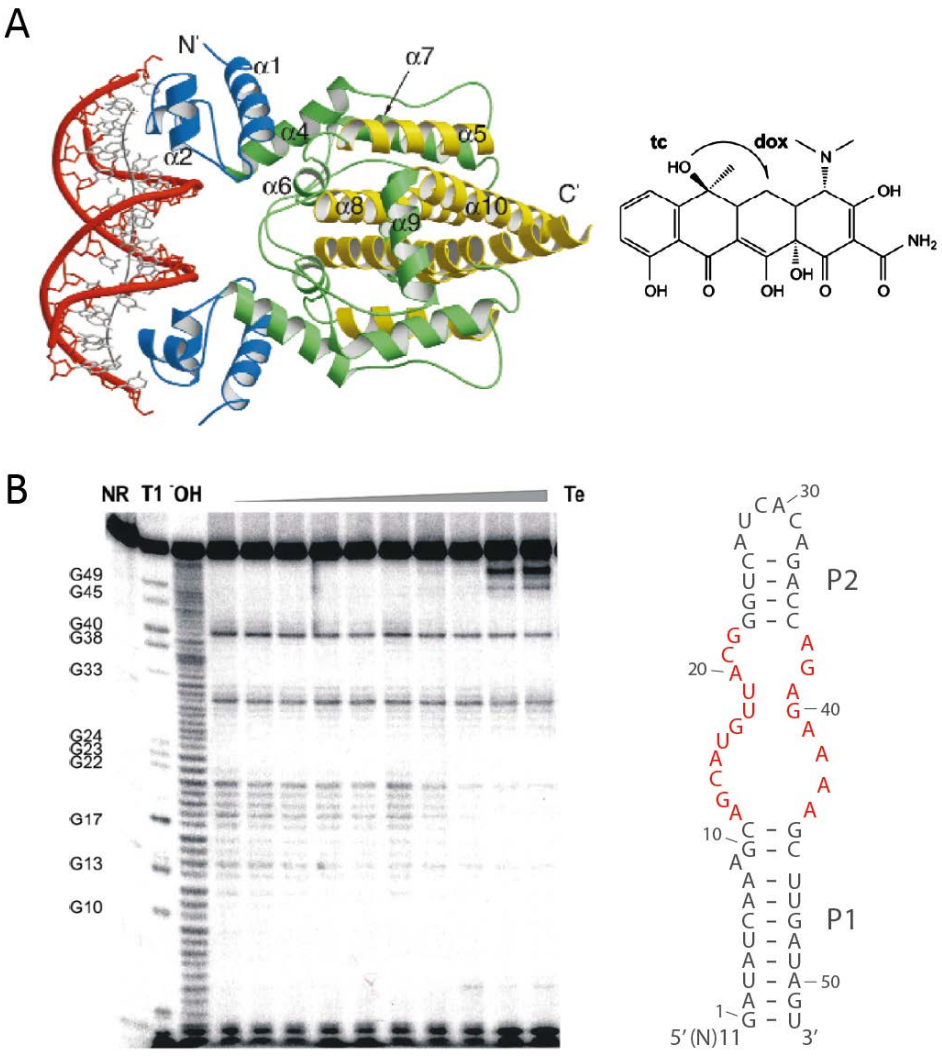


Figure 2.5 The tetracycline repressor (TetR) and the TetR-binding aptamer. (A) Crystal structure of the TetR-*tetO* complex. The two monomers are shown as yellow/green ribbons, two helix-turn-helix motifs are in blue and the DNA is displayed in red (adapted from [78]). (right) Structures of tetracycline and the derivate doxycycline (adapted from [83]). (B) In-line probing and the secondary structure of the TetR binding aptamer. Nucleotides involved in binding are shown in red (adapted from [82]).

The affinity of the TetR aptamer complex is extremely high and corresponds to that of the TetR bound to the *tetO* operator DNA [84]. Binding of tc or doxycycline (dox) to TetR leads to conformational changes within the protein resulting in DNA and RNA release, respectively, and consequently allows ligand-dependent reversible binding [85]. The applicability of the TetR aptamer system for the control of gene expression has been already confirmed by various approaches in different organisms. The aptamer was first used to activate TetR-controlled transcription in *E. coli* by displacing TetR from its DNA-binding site [82]. Next, portability and broader applicability of the system was documented with its successful use in the protozoon *Plasmodium falciparum* and in yeast [86, 87]. Further, an additional layer of regulation was added to the TetR aptamer system with the design of a theophylline-responsive TetR aptamer that proved functional [88]. Our group recently used the TetR aptamer to control miRNA biogenesis in human cells [68]. These approaches highlight the universal nature and potential of the TetR aptamer as a regulatory element.

2.3 pre-mRNA splicing

Most eukaryotic genes are expressed as precursor mRNAs (pre-mRNAs) that are converted to mRNA by splicing, an important step of gene expression in which introns must be removed and exons ligated together in order to produce correct mRNA. The accuracy of the splicing process involves the recognition of short sequences at 5' splice site (5SS) and 3' splice site (3SS) within the pre-mRNA that delimit the exon-intron boundaries [89]. The 3SS includes three sequence elements: the branch point sequence (BP), the downstream polypyrimidine (Py)-tract and the 3SS itself encoded by the dinucleotide AG. The splice sites sequences have shown a higher level of conservation in budding yeast compared to those in metazoans (Figure 2.6A). Introns are removed by two consecutive transesterification reactions (Figure 2.6B). First, the 2′ OH group of the BP adenosine carries out a nucleophilic attack on the 5SS. This splits the substrates into the upstream exon and the downstream RNA, in which the 5' of the intron is attached to the BP adenosine, forming a lariat structure. Second, the 3SS is attacked by the 3′ OH group of the 5′ exon, leading to the ligation of the 5′ and 3′ exons generating the mRNA and release of the intron. This process is catalysed by the spliceosome, and it is probably the most complex enzyme of the cell [90, 91].

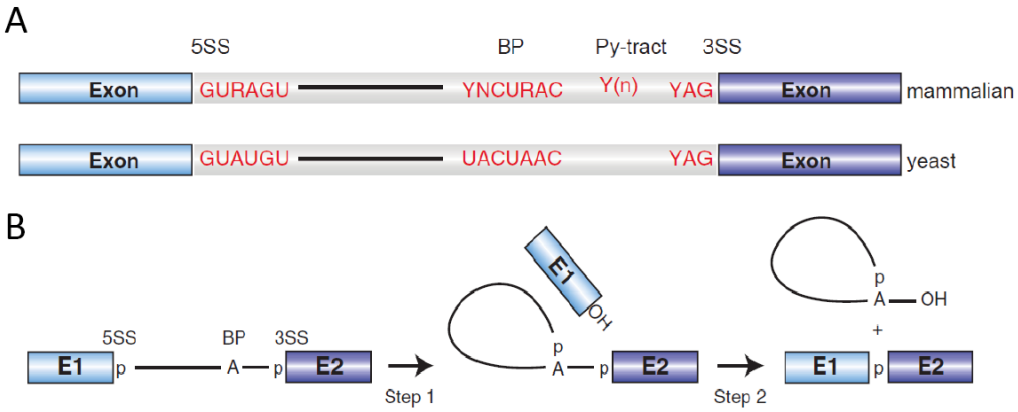


Figure 2.6 pre-mRNA splicing. (A) Conserved sequence elements of mammalian and yeast pre-mRNAs. Exons (in blue) separated by an intron (gray). The consensus sequences at the 5' splice site (5SS), the branch point sequence (BP), the polypyrimidine (Py)-tract and 3' splice site (3SS) are shown. N means any nucleotide, R is a purine, and Y is a pyrimidine. **(B)** Two steps of the transesterification reactions of pre-mRNA splicing (adapted from [89]).

The U2-dependent spliceosome contains the U1, U2, U5, and U4/U6 small nuclear ribonucleoproteins (snRNPs) and a large number of non-snRNP protein factors. Each snRNP is composed of a single uridine-rich small nuclear RNA (snRNA) and multiple proteins. Spliceosome assembly occurs by the ordered interaction of the spliceosomal snRNPs and numerous other splicing factors (Figure 2.7). Spliceosome assembly is highly dynamic in that complex rearrangements of RNA:RNA, RNA:protein, and protein:protein interactions take place within the spliceosome. For introns with the length of ~200-250 nt, the spliceosome initially assembles across the intron. In the earliest cross-intron spliceosomal complex (the E complex), the U1 is recruited to the 5SS and factors such as SF1/mBBP and U2AF interact with the branch point sequence (BP) and Py-tract, respectively. Next, the U2 stably associates with the BP, forming the A complex. The U4/U6·U5 tri-snRNP, which is pre-assembled from the U5 and U4/U6, is further recruited, generating the pre-catalytic B complex. Main rearrangements in RNA-RNA and RNA-protein interactions, leading to the destabilization of the U1 and U4, generate the activated spliceosome (the B^{act} complex). Next, catalytic activation by the DEAH-box RNA helicase Prp2, produces the B* complex, which catalyses the first of the two steps of splicing. This generates the C complex, which further catalyses the second step. The spliceosome then dissociates and, after some remodelling, the released snRNPs participate in other rounds of splicing [89, 90, 92].

Most mammalian genes contain multiple introns whose sizes differ from several hundred to several thousand nucleotides, whereas their exons have a more fixed

length of ~120 nt [93]. When intron length exceeds ~200-250 nt splicing complexes first form across an exon in a process called exon definition [94]. During exon definition, the U1 binds to the 5SS downstream of an exon and promotes the association of U2AF with the 3SS upstream of it which next leads to the recruitment of the U2 to the BP upstream of the exon. Splicing enhancer sequences within the exon recruit proteins of the SR protein family, which create a network of protein-protein interactions across the exon that stabilize the exon-defined complex. As the chemical steps of splicing occur across an intron, next from exon definition the 3SS must be paired across the adjacent intron with an upstream 5SS. The cross-exon interactions are first disrupted and the cross-exon complex is then converted into a cross-intron A complex, where a molecular bridge now forms between U2 and U1 bound to an upstream 5SS. This step is important in determining which 5′ and 3′ exon will finally be spliced. The switch from an exon-defined to intron-defined splicing complex is still poorly understood and probably leads to regulation of exon inclusion or skipping during alternative splicing evens [89, 95].

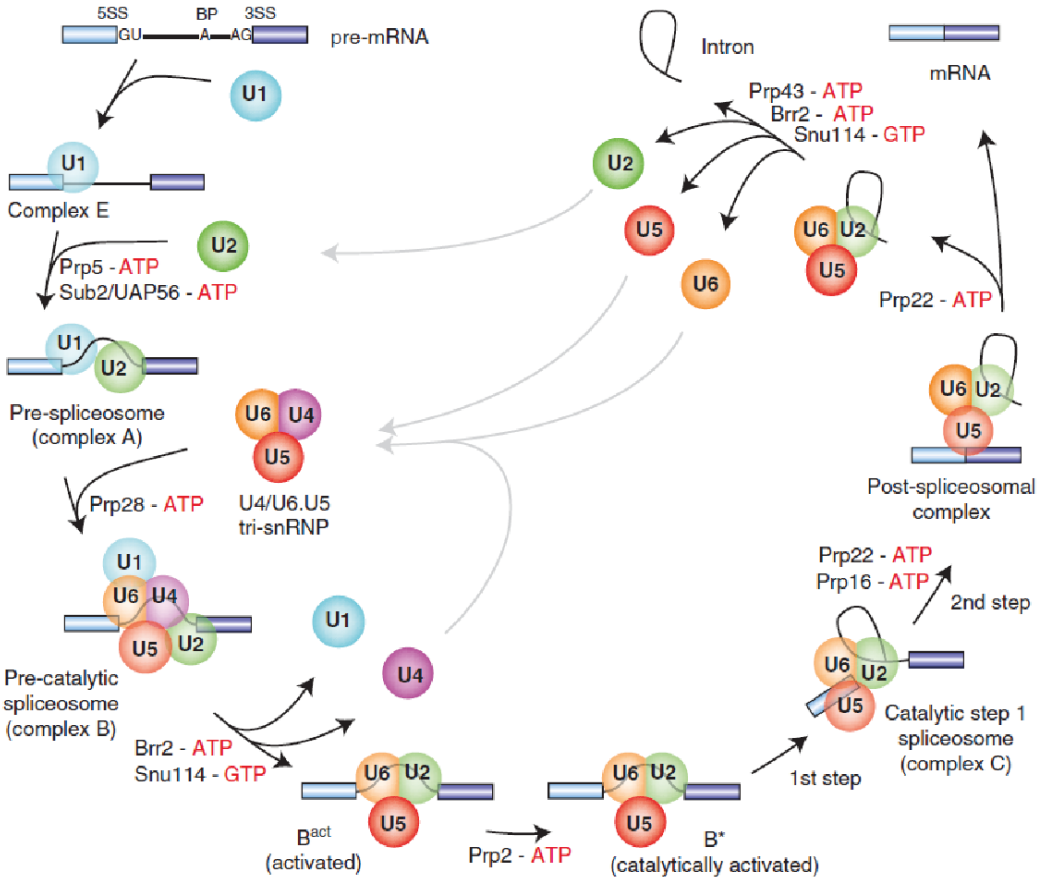


Figure 2.7 Pre-mRNA splicing circle. Pre-mRNA splicing is catalysed by the spliceosome, a multimegadalton ribonucleoprotein (RNP) complex comprised of five snRNPs (U1, U2, U5, U4/U6) and numerous proteins factors. Details description in the text (adapted from [89]).

2.3.1 Alternative splicing

Alternative splicing (AS) of pre-mRNA joins of different 5SS and 3SS, allowing one gene to express various mRNAs that encode proteins with diverse and even antagonistic functions [92]. AS in mammals is one of the most important cellular processes and mainly responsible for the diversity of the human proteome. Nearly 90% of the human genes are subjected to alternative splicing and disruption of the splicing machinery may lead to genetic diseases and cancer [96]. The main types of alternative splicing patterns (Figure 2.8) include the inclusion of alternative first and last exons, exon skipping (or cassette exon), intron retention, alternative 5SS and 3SS and more complex AS events that include mutually exclusive events, alternative transcription start sites and multiple polyadenylation sites. Exon skipping is the most common AS event in mammals. However, recent reports demonstrate that intron retention is also very frequently in human, occurring in nearly 75% of genes, and is a co- or post-transcriptional mechanism designed to reduce transcript levels during development [93, 97].

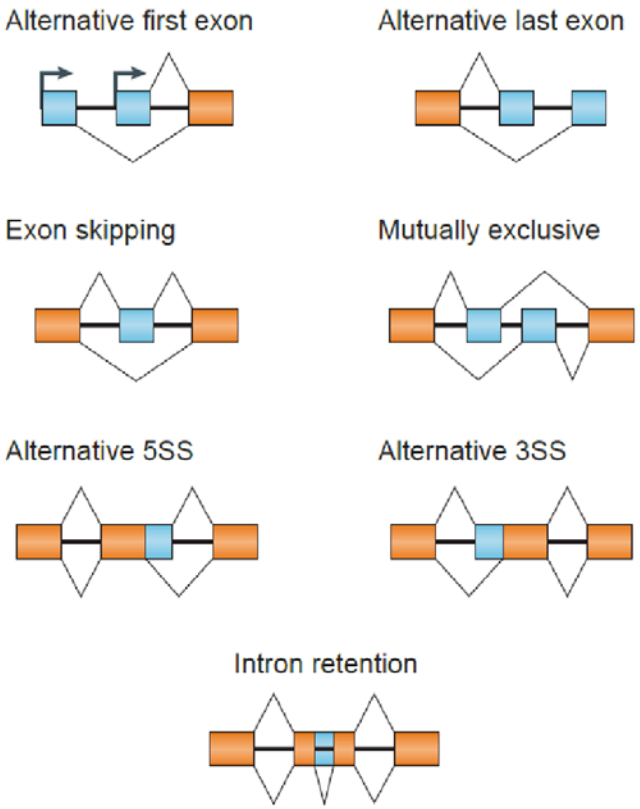


Figure 2.8 Types of alternative splicing. In all examples of AS the constitutive exons are shown in orange and the alternatively spliced regions in blue. Introns are represented as solid lines and thin lines indicate splicing events (adapted from [97]).

AS generates fragments of mRNA variability that can insert or remove amino acids, shift the open reading frame, or introduce a premature termination codon. Additionally, AS may influence gene expression by removing or inserting regulatory elements controlling translation, mRNA stability, or cellular localization. AS is strictly regulated in a cell type, developmental stage, gender or response to external stimuli manner. Splice site selection must therefore be precisely regulated in time and space. AS regulation is coordinated by intronic repressor and activator *cis*-elements distinct from the classical splicing sequences. Consequently, these elements serve to enhance or inhibit recognition of the classical splice sites by the core spliceosome. However, it is unclear why activation dominates in one cell type while repression dominates in another. Mutations that perturb this balance may disrupt AS and produce protein isoforms inappropriate for a cell type or developmental stage [97, 98]. Full understanding of alternative and constitutive splice sites selection is complicated as features of regulatory sequences and splice site signals are often short and degenerate. Moreover, the best bioinformatic tools are only 50% precise in predicting actual splice sites over frequent and similar sequences that are not used. About 50% of the documented mutations result in aberrant splicing. Some have been reported to disrupt splicing by creating a new splice site or by interfering with splicing regulatory elements. Splicing mutations then may play a more important role than previously though in human hereditary disease and cancer [99, 100, 101, 102]. Reprogramming of aberrant splicing could provide a novel approach for the development of gene therapies to tackle disease phenotypes. For this purpose, it is necessary to engineer tools that allow precise and timely control of the splicing process.

2.3.2 Splicing control with aptamers

The only documented natural riboswitches found in eukaryotes are thiamine pyrophosphate riboswitches in plants and filamentous fungi [55,56,57,39]. Eukaryotic riboswitches regulate gene expression through AS (Figure 4.2B). In these examples the aptamer masks a 5SS, causing an alternative 5SS selection. In the presence of a ligand the 5SS is unmasked, causing inclusion of an alternative exon which consequently leads to translation of an upstream open reading frame or inclusion of a premature stop codon. In higher plants, the aptamer located in the 3' UTR sequesters both 5 and 3SS, leading to inclusion of a polyadenylation sequence and a mature RNA

transcript and in the presence of a ligand both SS are unmasked, leading to removal of the polyadenylation sequence and consequently a transcript degradation.

Splicing regulation by RNA-based systems has been demonstrated in yeast and mammalian cells. Suess group has shown an inhibition of pre-mRNA splicing mediated by the tc aptamer in yeast [60]. The tc aptamer is inserted in a way that the consensus sequence of 5SS is integrated within the closing stem of the aptamer. The 5SS is recognized by the U1 only in the absence of tc. A regulatory factor was achieved up to 16-fold. It is a highly efficient aptamer-based conditional gene regulation system in yeast and attempts to transfer it into human cell lines failed so far. Gaur and co-workers have shown splicing inhibition through theophylline-sequestering of branch point sequence and 3SS [58, 59]. In an *in vitro* splicing assay a 4-fold reduction of gene expression is observed upon addition of theophylline. In HeLa cells the effect was not strong enough to influence a switch in splicing site choice. Smolke group has used MS2 system to control cell fate through exon skipping to regulate the expression of herpes simplex virus-thymidine kinase in human cells [74]. Currently, an exon skipping control with tc-aptamer by blocking a 3SS was demonstrated [61]. However, the regulatory effect of these devices was modest (~2- to 4-fold). Despite the extraordinary importance of AS for gene regulation, the number of synthetic splicing devices is fairly limited and suggests that the full potential for their development has been far from realised to date.

3 Scope

The field of synthetic biology strives to apply engineering principles to the design and build-up of biological systems for defined purposes. At the heart of all of synthetic biology applications is the precise control of gene expression – turning the right gene on at the right time and at the right level. The RNA-based gene regulators constitute a fundamental component of the synthetic biology ‘toolbox’. Combination of versatility, designability and easy characterization makes RNA the most powerful substrate for the engineering of gene expression within cells. To date, the most substantial progress in functionality of RNA-based devices has been demonstrated in microorganisms, while even the best mammalian systems are burdened with lower gene regulation efficiency than their bacterial counterparts, necessitating the development of tools specific to these types of cells. Precisely tuned human gene expression control with synthetic RNA regulators may enable a wide range of practical applications in the realm of life sciences.

Therefore, the focus of the studies undertaken in course of preparation of my doctoral thesis was the design and development of versatile, robust and reversible RNA-based devices active in human cells through engineering and exploiting the potential of the TetR aptamer as a key synthetic element. Three strategies to use the TetR aptamer to control protein output by intron retention, alternative 3’ splice sites and regulation of translation have been explored. The following sections of the hereby presented thesis describe application and universality of the TetR aptamer in human synthetic biology.

4 Results

4.1 Project I: Universal splicing module based on TetR aptamer to control gene expression in human cells

Fine-tuning of gene expression is desirable for a wide range of applications in synthetic biology and medicine and to this purpose RNA regulatory devices can be applied [8, 17, 18, 19]. Splicing of pre-mRNAs is an essential process in mammalian cells that generates a highly diverse proteome through networks of coordinated splicing events. Reprogramming of aberrant splicing could provide a novel approach for the development of gene therapies to tackle disease phenotypes. Towards this purpose, it is necessary to engineer tools that allow precise and timely control of the splicing process. Despite the extraordinary importance of alternative splicing for gene regulation, the number of synthetic splicing devices is fairly low [74, 59, 61] and suggests that the full potential for their development has been far from realized to date. The aim of this project was to designed a versatile and highly efficient TetR Splicing Device (TSD) for controlling gene expression in human cells that makes use of an RNA aptamer that is recognized by the TetR aptamer [82]. Portability of the splicing device was shown through its functionality in different reporters and endogenous gene context.

4.1.1 Results

The 53 nt long TetR aptamer was used to set up a switching device to control intron retention. The overall idea was to place the aptamer close to the 5SS in a way that binding of TetR to the aptamer would efficiently inhibit splice site recognition. Consequently, the intron is retained. Addition of dox then releases TetR from the aptamer freeing the 5SS resulting in correct splicing (Figure 4.1A). We made use of a previously established system for the firefly luciferase (FLuc) gene [103]. A chimeric intron (CI) composed of the 5SS from the first intron of human *β*-globin gene and 3SS from the intron of an immunoglobulin gene heavy chain variable region was introduced into the FLuc. The CI intron was designed in a way that FLuc activity only occurs when the mRNA is correctly spliced. Unspliced mRNA should not be exported. If it does escape from the nucleus, premature stop codons in every reading frame in the intron sequence would either lead to rapid mRNA degradation via the nonsense-mediated decay pathway or to translation of a truncated protein.

First, the influence of the position of the intron on FLuc expression was examined. Thirteen different positions mimicking exonic splice features were selected within the FLuc cDNA, which resulted in the constructs CI1-13. Next, to analyse if the TetR aptamer can influence 5SS recognition, the aptamer was inserted into the intron directly behind the 5SS, resulting in the constructs CI1T-CI13T. HEK293 cells were transiently transfected with these constructs together with a plasmid expressing TetR. TetR itself was expressed from a strong CMV promoter and modified at its N-terminus with a nuclear localisation signal (NLS) from c-MYC that significantly increased its localisation in the nucleus. The insertion of the TetR aptamer into CI influenced FLuc activity to a different degree. The expression of TetR lead to significant reduction of FLuc activity in most constructs, yet upon addition of dox FLuc activity was completely restored (Figure 4.1B). Only three out 13 tested positions (CI4T, CI5T and CI12T) exhibited a significant dynamic range of regulation (8.3-, 14.7- and 10.2-fold, respectively). These constructs have both a high dynamic range but also a low basal activity in the OFF state. Unfortunately, the constructs CI5T and CI12T also showed a very low overall expression level in the ON state. Therefore, the construct CI4T was selected for further characterisation. TetR aptamers placed 9 nt (or more) downstream of the 5SS had no influence, probably due to free accessibility of the splice site for U1. It was also modified the length and the stability of the aptamer closing stem. Only stem stabilities between -11 and -16 kcal/mol allow efficient regulation.

From all tested constructs the CI4T5 turned out to be a splice switch with an 8-fold dynamic range (Figure 4.1C). Further, the introduction of the two mutations A15U and A20C that destroy TetR binding lead to a complete loss of regulation, indicating that switching is indeed mediated by TetR aptamer interaction (Figure 4.1D). Moreover, the system up to 90% with only 0.7 µM dox was activated and it was confirmed that dox has no influence on cell viability at the concentration used. Finally, it was demonstrated that TetR can also influence the responsiveness of the system in a dosage-dependent manner [103].

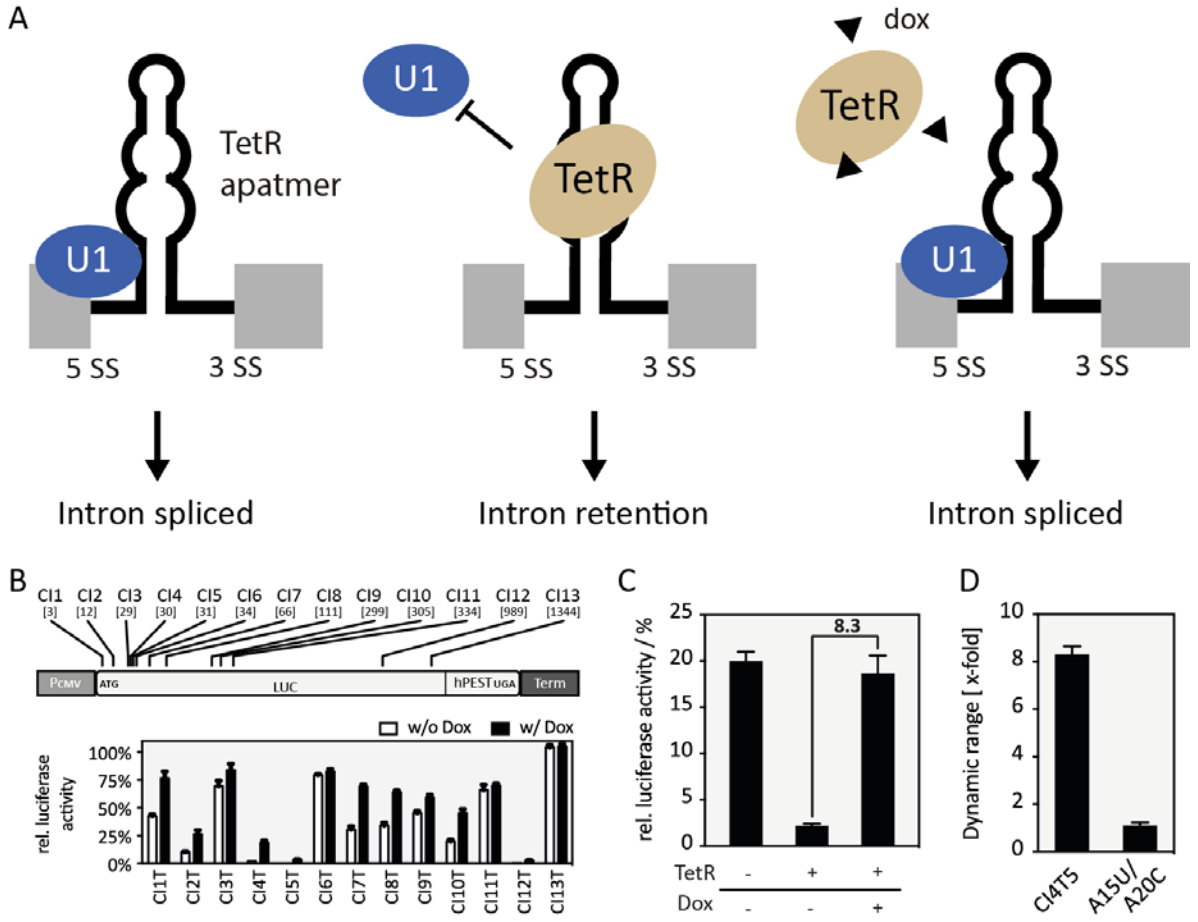


Figure 4.1 TetR aptamer controls intron retention. (A) Schema of the proposed model. (left) In absence of TetR (in yellow), U1 (in blue) can recognize the 5SS, the intron will be spliced out. (middle) In the presence of TetR, U1 cannot recognize the 5SS, consequently the intron is retained. (right) Splicing can be restored by addition of dox that leads to conformational changes of TetR and the release of the aptamer. Exons are displayed as boxes and intron with TetR aptamer as a line. **(B)** Displayed are the positions of the intron with TetR aptamer insertions (positions shown in nucleotides relative to the ATG start codon) and their effect on FLuc activity. Constructs CI1-13T were co-transfected with plasmid expressing TetR. FLuc activity was measured 24 h after transfection in the absence or presence of 50 µM dox. Shown are the fluorescence values without (white bars) and with dox (black bars). Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means from three independent experiments. **(C)** Displayed is FLuc activity of the CI4T5 switch in all three states. The regulatory activity for CI4T5 is shown. CI4T5 was co-transfected with TetR and FLuc activity was measured 24 h after transfection in the absence or presence of 50 µM dox. **(D)** Mutations A15U and A20C in the aptamer destroy TetR protein binding, which leads to loss of switching behaviour.

Under cellular conditions, synthetic regulatory devices are influenced by the sequence environment and/or additional cellular factors, also described as intrinsic and extrinsic noise. Therefore, it is always challenging to transfer regulatory devices from one context into another [104,105,106]. The integration of the CI4T5 from FLuc into a similar position of a *gfp* gene resulted in only 2-fold regulation [data not shown]. It contributed further to the CI4T5 transformation into a splicing device (TSD - TetR splicing device)

that can be applied to any *gene of interest* (*GOI*) with constant efficiency irrespective of its (genomic) context.

First, a minimal sequence of CI4T5 important for efficient switching was defined. Due to the short size, the complete first exon (27 nt) from CI4T5 together with the intron with the TetR aptamer and six nucleotides from the second exon was taken and placed it in front of a *gfp* gene (Figure 4.2A). HEK293 cells were transiently transfected with this construct (named SP) together with a plasmid expressing a TetR::mCherry fusion protein. GFP and mCherry fluorescence were monitored 24 h after transfection in the absence or presence of dox using flow cytometry. mCherry expression was used to normalize for variation in transfection efficiency. The construct exhibits a similar expression level as the luciferase at the ON and OFF state (18% and 3%, respectively), as well as dynamic range of 6.6 (Figure 4.B).

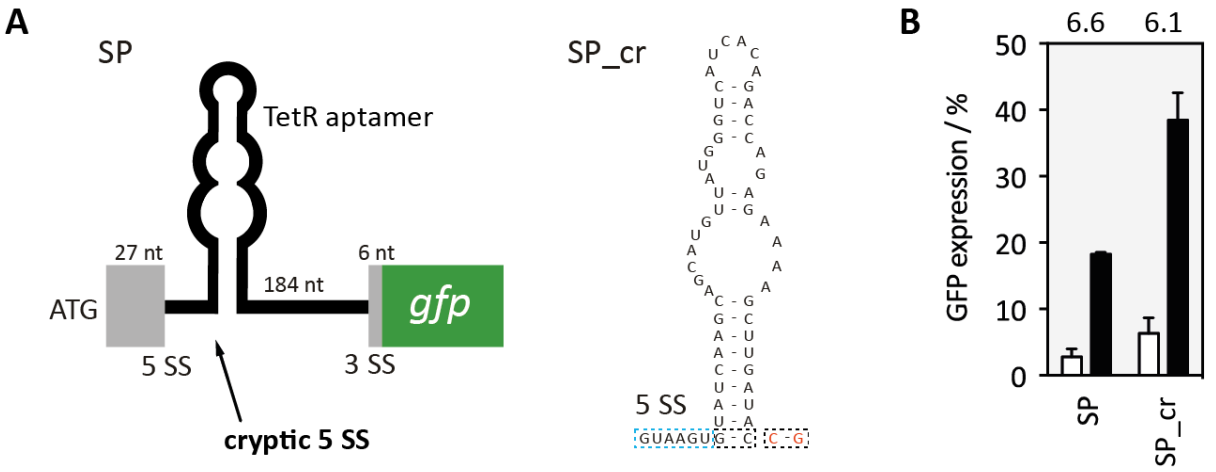


Figure 4.2 Defining the switching sequence and influence of cryptic 5SS within CI4T5. (A) The switching part was placed in front of *gfp* and SP_cr construct with mutated cryptic 5SS, G37-C87 to C37-G87. Introduced mutations are marked in red and 5SS is shown in blue. **(B)** Displayed are the GFP fluorescence SP and SP_cr constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. Constructs were co-transfected with TetR:mCherry. GFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry and mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

Next, a cryptic 5SS in the stem of the aptamer was identified that could be responsible for the overall expression level being so low. This site was mutated (G3C, C53G) generating the construct SP_cr. The new construct exhibited two times higher GFP expression with a dynamic range nearly remaining unchanged (Figure 4.2B). Additionally, it was assessed whether and to what extent the stability of the aptamer

influences regulation. Two constructs were designed, one with shorter and second with a longer closing stems (SP_cr1-3, Figure 4.3A). Whereas destabilization of the stem decreased the dynamic range, a stabilization of the stem resulted in improved switching activity (8.3- compared to 6.1-fold) and higher GFP expression of the ON state compared to the SP_cr (Figure 4.3B).

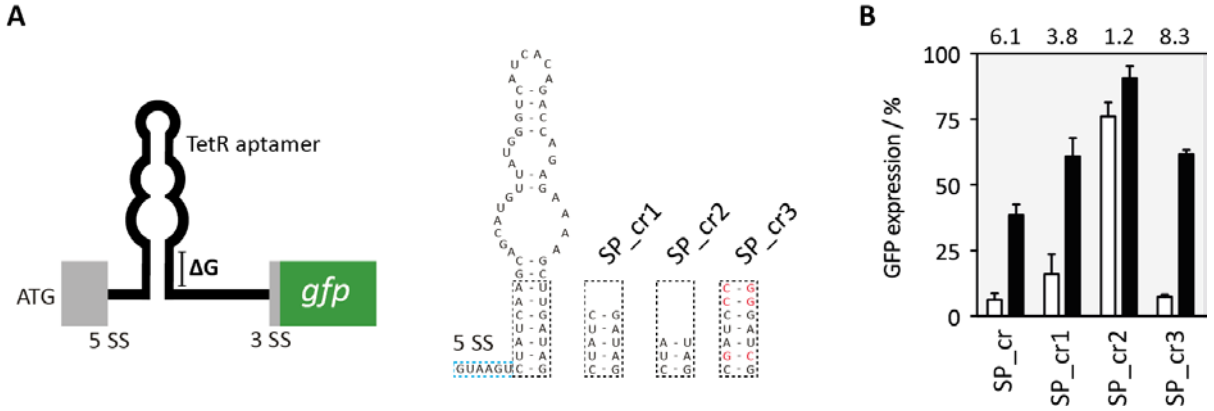


Figure 4.3 Stem modifications within SP_cr. (A) The stem was truncated of 2 or 4nt, generating constructs SP_cr1 and SP_cr2, respectively. Additionally, the stem was stabilized cloning SP_cr3 construct. Introduced mutations are marked in red and 5SS is shown in blue. **(B)** Displayed are the GFP fluorescence for all constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. Constructs were co-transfected with TetR:mCherry. GFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry and mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

However, the GFP expression of the construct is still only 50% compared to the wild type GFP expression. We speculated that the weak intron removal can be due to exonic splicing silencers (ESS) located in the first exon. RT-PCR data indicates an accumulation of pre-mRNA and low level of mRNA (Figure 4.4A). The sequence of the first exon was scanned using SpliceAid2, a database of human splicing factors and RNA target motifs. The program predicted the sequence AAGAAGGGC as a putative binding motif for heterogeneous nuclear ribonucleoprotein (hnRNP) H1, H2, H3 and F known as ESS [107]. The proposed ESS motif was mutated (SP_cr4-6, Figure 4.4B). All constructs exhibit significantly higher GFP expression, both at the OFF and ON state, but a decrease in the dynamic range compared to SP_cr (Figure 4.4C). Additionally, the construct SP_cr7 was cloned carrying an additional ESS motif. The GGAAGAAC motif was chosen that is recognized by the strong exonic splicing enhancer (ESE) SRSF1 [108] (Serine/Arginine-rich splicing factor 1). Here the construct exhibits 2.5-fold

dynamic range while up to 100% intron is spliced in the present of dox (Figure 4.4C). However none of the constructs showed increased regulation.

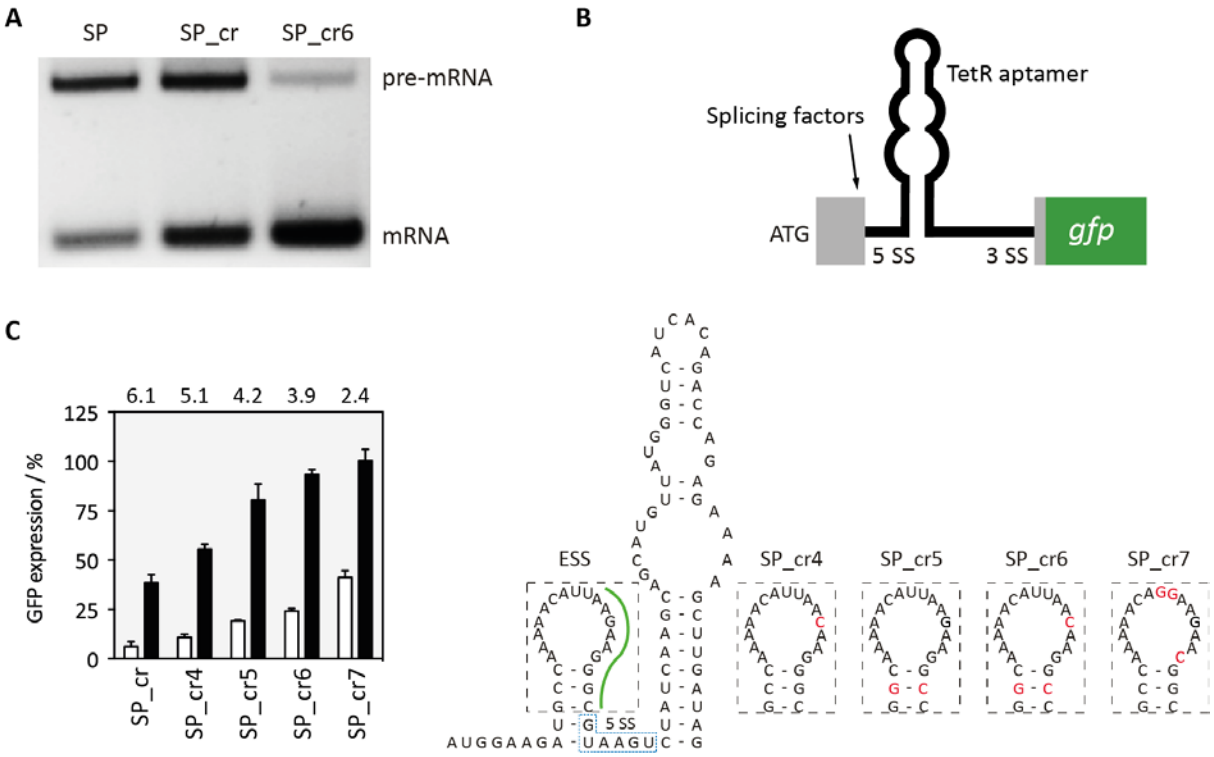


Figure 4.4 Mutated putative splicing motif. (A) Splicing pattern of SP, SP_cr and SP_cr6 constructs visualised by RT-PCR. Cells were transiently transfected with these constructs. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. Upper bound corresponds to pre-mRNA and lower to mRNA. The spliced products were cloned using CloneJET PCR cloning kit and sequenced for verification. Experiment was repeated three times with similar results. **(B)** Using SpliceAid2, program predict the AAGAAGGGC sequence (green line) upstream 5SS, as a putative binding motif for exonic splicing silencers (ESS) like hnRNP H1, hnRNP H2, hnRNP H3 and hnRNP F. These ESS belong to the hnRNP family characterized as mediators of splicing silencing. Constructs SP_cr4-6 with mutated ESS motif were cloned. Additionally, a construct SP_cr7 that carries an exonic splicing enhancer motif (GGAAGAAC), recognized by strong splicing enhancer SRSF1, was cloned. Introduced mutations are marked in red and 5SS is shown in blue. An additional folding, upstream the aptamer structure, was detected by a RNA structure prediction programs (mfold). **(C)** Displayed are the GFP fluorescence for all constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. Constructs were co-transfected with TetR:mCherry. GFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry and mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

With the construct SP_cr used, 39 additional nucleotides remain after removal of the intron at the 5' end of the open reading frame. This corresponds to 13 amino acids attached to the N-terminus of the protein of interest. This may not be tolerated by several proteins. The initial attempt to eliminate the additional nucleotides by moving

the start codon towards the 5SS (SP_cr8-10) was unsuccessful. Truncation of the exon to 12 nucleotides (SP_cr11) also showed negative results (Figure 4.5).

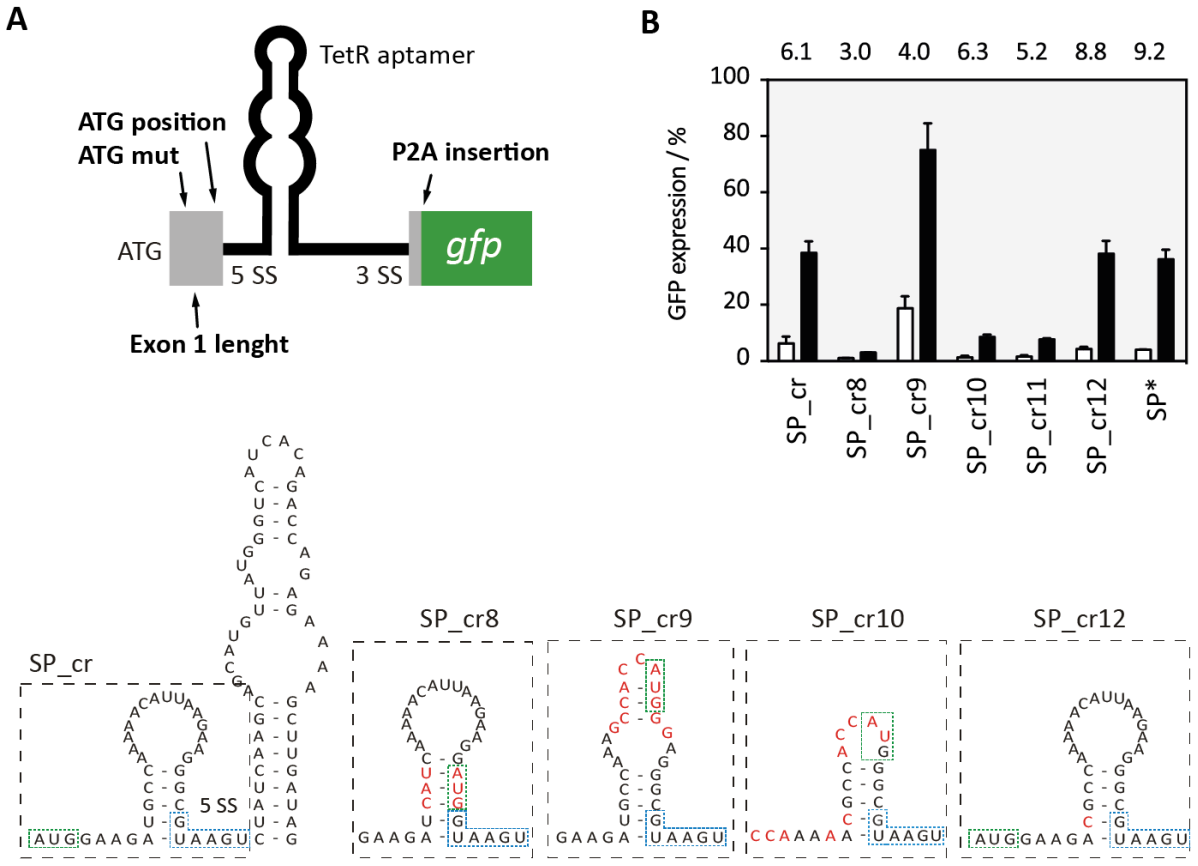


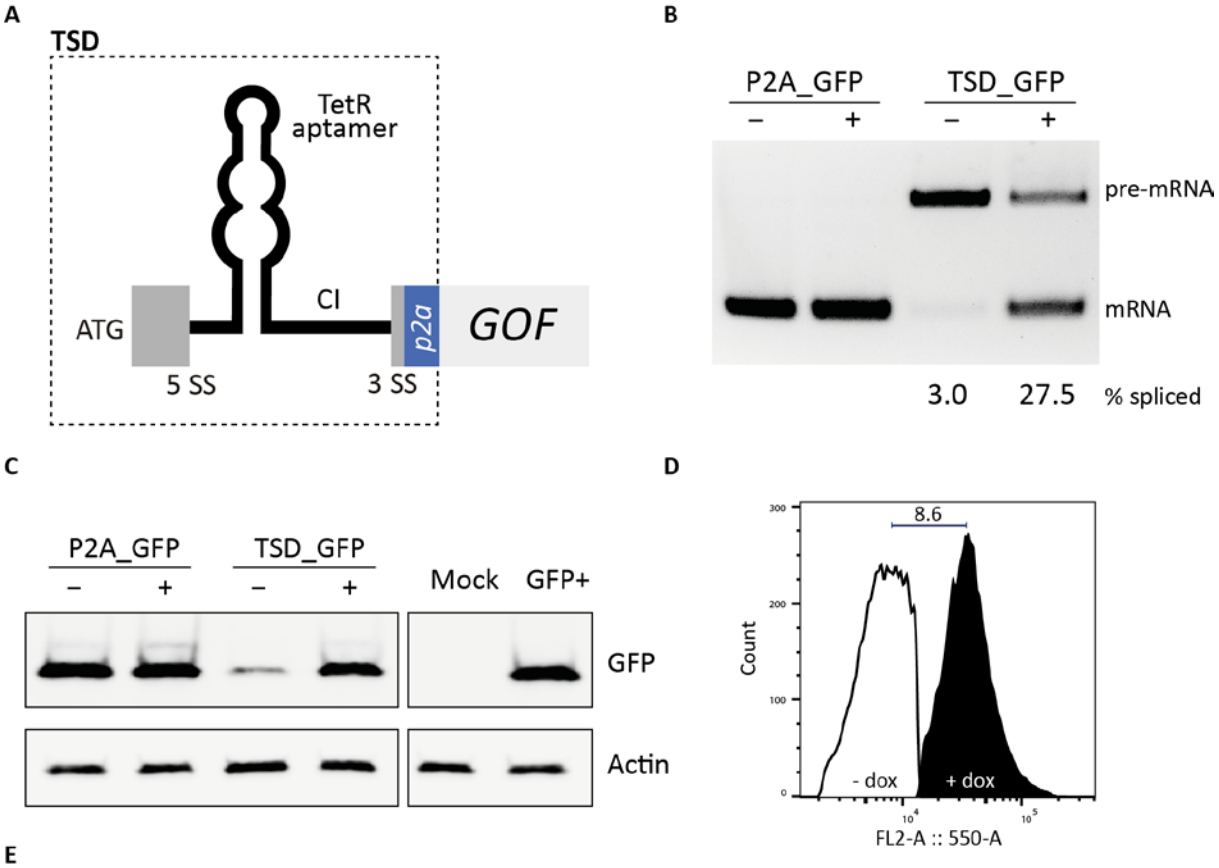
Figure 4.5 Reduction of the aptamer-born amino acids. (A) Scheme of cr_5ss modifications. SP_cr8-10 are constructs with shifted ATG position at the end of exon 1. For constructs SP_cr9 and SP_cr10 additional mutations were inserted within first exon to introduce Kozak sequence to enhance translation. Introduced mutations are marked in red, 5SS is shown in blue and ATG start codon in green. Construct SP_cr11 has truncated first exon to 12 nt. In the construct SP_cr12, an additional ATG in the first exon that is out of frame was mutated (U9C, positions relative to the ATG start codon) to keep only one codon start. The porcine teschovirus-1 2A (P2A) peptide was introduced into SP_cr12, generating SP*. **(B)** Displayed are the GFP fluorescence for all constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. Constructs were co-transfected with TetR:mCherry. GFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry and mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

Additionally, a start codon located in the first exon was removed, resulting in an improvement to 8.8-fold (SP_cr12). Finally, we inserted a 'self-cleaving' peptide between the device and the start codon to remove the additional exon-born amino acids (construct SP*). The porcine teschovirus-1 2A (P2A) peptide was used with the sequence GSGATNFSLLKQAGDVEENPGP. The cleavage site of the peptide is

located between its last two amino acids [109, 110]. Consequently, only one proline is added to the N-terminus of the protein of interest. It is expected that the majority of proteins will tolerate this minimal invasive change in their protein sequence. The addition of P2A domain slightly decreased the *gfp* expression, but increased the dynamic range to 9.2-fold (Figure 4.5B).

Next, SP* as the best construct was chosen as TSD_GFP. The sequence of the complete device (292 nt) that can be transferred upstream of the coding sequence of any gene of interest is given in Figure 4.6. Further, a careful characterization of TSD_GFP was carried out. The switching capability at the mRNA level was analysed (Figure 4.6B). Intron retention was visualized by RT PCR using oligonucleotides binding to exon 1 and 2 and quantified by qPCR. In addition, the protein level was detected by western blot (Figure 4.6C). The qPCR data are consistent with GFP expression measured by flow cytometry (Figure 4.1.5B) supported by the Western blot experiment. Additionally, TSD_GFP was stably integrated using the HF1-3 Flp-In system. Stably integrated TSD_GFP shows the same regulation as in the transient situation (Figure 4.6D). Moreover, it was shown that TSD_GFP was functional in HEK293, HeLa, A549 and CHO cell lines [data not shown].

Finally, portability of the TSD in the context of other reporter gene mCherry and the human transcription factor MAX was assessed. Analogue as in the GFP context, the TSD was inserted at the 5’ end of the coding sequence of both genes, generating TSD_mCherry and TSD_MAX-GFP, respectively. The constructs were co-transfected with the plasmid expressing TetR. mCherry and MAX-GFP expression were measured 24 h after transfection in the absence or presence of 50 µM dox using flow cytometry. In the case of mCherry control TSD was first inserted in front of mCherry, generating TSD_mCherry-1. This construct exhibits only ~1.6 regulation fold (Figure 4.7A). In the absence of dox the construct shows high mCherry expression. To check if the TetR aptamer can control intron retention event, RT-PCR was perform. The RT-PCR clearly shows that TetR aptamer leads to intron retention in the absence of dox (Figure 4.7B), and mCherry expression should be reduced at OFF state as mRNA transcript produced in the case of intron retention state contain introduced premature stop codon. This transcript should be degraded by a nonsense-mediated mRNA decay pathway.



E

ATGGAAGACGCCAAAAACATTAAGAAGGGCGTAAGTCTATCAAGCAGCATGTTATGGGTCATCACAGACCAGAGAAAAGCTTG
ATAGATCAAGGTTACAAGACAGGCGTACGGAGACCAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATA
GGCACCTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGCCAGCGACGCGTGGAAGCGGAGCTACTAACTTCAGCCTG
CTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCT

Figure 4.6 Validation of TSD. (A) General sketch of the TSD design. The TSD element is highlighted by a dotted line inserted in front of a gene of interest (GOI). Exons are displayed as boxes and the intron as a line, the self-cleaving domain P2A is displayed as a blue box. **(B)** Splicing pattern of P2A_GFP (control without the aptamer) and TSD_GFP (before SP*) constructs visualised by RT-PCR. Cells were transiently co-transfected with these constructs and with plasmid expressing TetR and treated with (+) or without (-) 50 µM dox for 24h. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. Upper bound corresponds to a pre-mRNA and lower to mRNA. Quantification of % spliced was determined by qPCR analysis. **(C)** The protein expression of P2A_GFP and TSD_GFP constructs was analysed by Western blot analysis. Anti-GFP was used for GFP expression and anti-Actin was used as a loading control. Mock was transfected with TetR::mCherry and GFP+ is GFP control without P2A domain. Experiments were repeated three times with similar results. **(D)** Additionally, P2A_GFP and TSD_GFP constructs were stably integrated into HF1-3 cell line with Flp-In system. Generated cell lines expressing these construct were subsequently transiently transfected with TetR::mCherry and treated with or without 50 µM dox for 72h. The histogram for TSD_GFP with the regulatory activity is shown. GFP and mCherry was excited using a 488 nm laser with a 550/30 filter and 561 nm laser with a 610/20 filter, respectively. Populations were selected by gating out the GFP background signal of untransfected cells with TetR::mCherry by FlowJo software. **(E)** The TSD sequence. Colour coding: exon sequences highlighted in grey, TetR aptamer inside the intron in yellow and P2A self-cleavage sequence in blue.

Therefore, another AUG is potentially use by ribosome which leads to correct mCherry production. The switching module may mimic upstream ORF (uORF) due to the short sequence, and another AUG in frame within mCherry sequence may be used. Several described mechanisms enable expression of the downstream ORF despite the presence of an uORF. Ribosomes may bypass the uAUG codon. The context of surrounding nucleotides is one determinant of AUG codon recognition by the ribosome. However, many uAUG codons do not have the optimal sequence, leading to the prediction that some scanning ribosomes are likely to bypass them, a process known as 'leaky scanning'. Even if the ribosome recognizes an uAUG codon and translates the uORF, it might reinitiate at a downstream AUG codon [111, 112, 113]. At the beginning of mCherry sequence the AUG codon in frame (it is 24 nt after TSD) is present, that may produce an active mCherry. To test if another AUG can be used by ribosome, a construct TSD_mCherry with mutated 26G to 26A within mCherry sequence was cloned. The TSD_mCherry construct exhibits ~10-fold regulation (Figure 4.7A) demonstrating that TSD can control mCherry gene and that 'leaky mechanism' may have happened. Therefore, it will be necessary to scan a sequence of GOI for a presence of putative AUG codons and mutate them.

While, the MAX gene was fused with GFP as a reporter at its C-terminus to allow detection of MAX protein expression by flow cytometry. MAX-GFP construct exhibits a 5.4-fold regulation (Figure 4.7C). The dynamic range of MAX-GFP was slightly lower compared to the TSD_GFP or TSD_mCherry constructs possibly due to the fusion of MAX with GFP.

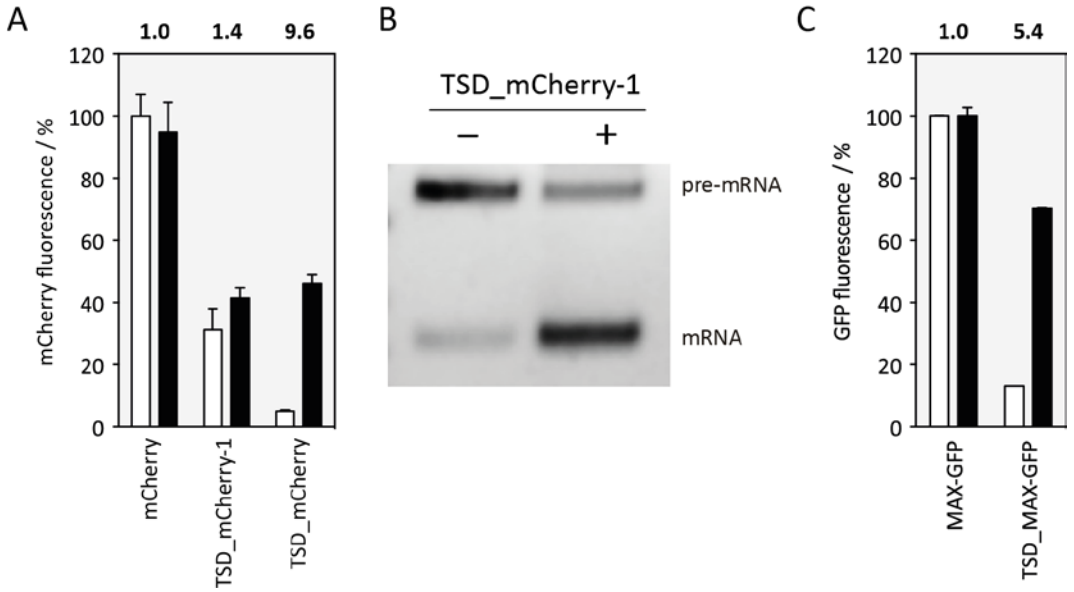


Figure 4.7 TSD portability. (A) TSD controlling mCherry expression. TSD_mCherry-1 and TSD_mCherry constructs carry TSD placed in front of mCherry coding sequence. Additionally, TSD_mCherry has mutated 26G to 26A within mCherry sequence. Control gene has inserted coding sequence from TSD (without TetR-intron sequence). Constructs were co-transfected with TetR. mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry. Displayed are the mCherry fluorescence values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. **(B)** Splicing pattern of TSD_mCHerry-1 construct visualised by RT-PCR. Cells were transiently co-transfected with this construct and with plasmid expressing TetR and treated with (+) or without (-) 50 µM dox for 24h. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. Upper bound corresponds to a pre-mRNA and lower to mRNA. **(C)** TSD in the context of MAX transcription factor. MAX was additionally fused with GFP as a reporter at its C-terminus to allow detection of MAX expression by flow cytometry. Constructs were co-transfected with the mCherry tagged TetR. mCherry and GFP expression was measured 24 h after transfection in the absence or presence of 50 µM dox. Displayed are the GFP fluorescence values without (white bars) and with dox (black bars) and the regulatory activity above each construct are shown. (A and C) Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

In summary, these results clearly demonstrate that TSD was able to control the expression of several genes of interest, with splicing regulation directly translating into protein output which additionally highlights the genetic modularity and robustness of the device.

4.2 Project II: Inducible control of nuclear import using TetR aptamer

The nucleo-cytoplasmic transport is essential for cell physiology and aberrant spatiotemporal localization of proteins leads to various forms of cancer or other diseases [114, 115]. The manipulation of protein localisation by targeting the protein folding, signal transduction and nuclear transport would provide a promising antiviral and anticancer strategy [116]. However, despite the importance of this process, this has not yet been extensively explored. Currently, few inhibitors have been identified to interfere with the nuclear transport [117, 118]. Therefore, development of synthetic devices that allow spatial regulation of gene expression would be valuable in phenotypic studies of essential genes and therapeutic strategies that target protein localization [116, 119]. Controlling nuclear localisation as a strategy to control protein activity can be accomplished at the mRNA or protein level. Previously, two tools were engineered for mRNA transport, however, these systems are only capable of cytoplasmic redistribution of mRNA in the cells, one of which is limited to yeast [120, 121]. Other devices able to control mRNA shuttling between cytoplasm and nucleus in human cells have not been demonstrated. All existing tools controls protein activity at the level of protein localisation, which are based on chemical [120, 122, 123, 124] or light-inducible systems [119, 125, 126]. In the second part of the thesis we created the first inducible model for alternative 3' splice site (A3SS) recognition with the TetR aptamer that leads to production of splice variants with different subcellular localization. The practicality of the system was further shown by controlling nuclear import in human cells, which may provide an alternative for therapeutic strategies.

4.2.1 Results

As a first step toward creating a device for the control of A3SS using the TetR aptamer system, the MINX intron with flanking exon sequences was placed in front of an *egfp* reporter gene. This small intron derives from the adenovirus and is efficiently spliced in human cells [127] (Figure 4.8). Next, the TetR aptamer was placed upstream of the canonical 3SS (Figure 4.9A). The TetR aptamer contains five potential sequences that can be recognized as A3SS (A3SS-T1). In the first step, two CAGs in the aptamer loop were mutated (A3SS-T2; Figure 4.9B). The AAG should not be used as 3SS as it is a weak splice site compared to the upstream CAGs. GAG is generally not used as a 3SS in eukaryotic cells with few exceptions [128].

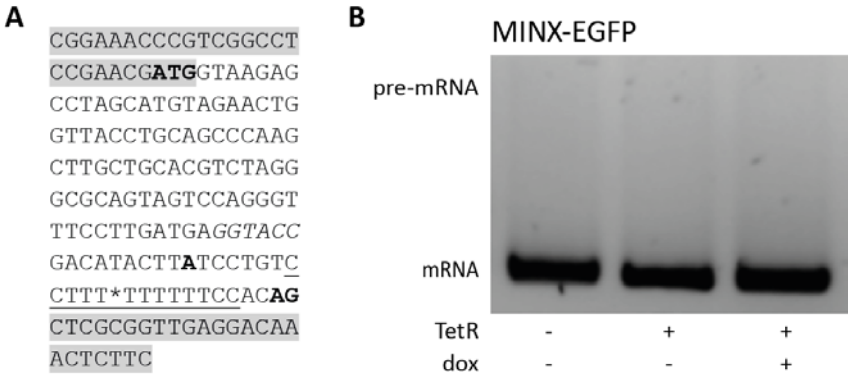


Figure 4.8 MINX intron. (A) sequence of used MINX intron. Flanking exons are marked in grey. Start ATG codon, branch point adenosine and AG are in bold and the Py-tract is highlighted. Introduced KpnI restriction site is in italic. Position of aptamer insertion indicated by *. **(B)** Splicing pattern of control MINX-EGFP construct visualised by RT-PCR. HeLa cells were transiently transfected with the MINX-EGFP, transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. The bound corresponds to spliced product. Experiment was repeated three times with similar results.

HeLa cells were transiently transfected with these constructs together with a plasmid expressing TetR and treated with or without 50 µM dox for twenty-four hours. TetR itself was expressed by a strong CMV promoter and modified at its N-terminus with a nuclear localization signal (NLS) from c-MYC that significantly increased its localization in the nucleus. Splicing pattern was visualized by RT-PCR using primer pairs binding to both exons. In both constructs spliceosome recognized only the first A3SS in the aptamer and the distal 3SS (Figure 4.9B). Interestingly, for A3SS-T1 construct enhanced A3SS usage in the present of TetR is observed.

Next, based on A3SS-T1 construct, the aptamer stem length and composition (A3SS-T3-T7; Figure 4.10), as well as the polypyrimidine(Py)-tract length and position relative to the aptamer was characterized and optimized (A3SS-T8-T13; Figure 4.11). The short aptamer stem in A3SS-T3, as well as more pyrimidines upstream the A3SS in A3SS-T5 leads to the exclusive recognition of the A3SS. The aptamer stem stabilization by additional G-C base pairing in A3SS-T6 enhanced switching behaviour most prominently in the presents of dox, as compared to A3SS-T1. Also, the Py-tract position and its composition relative to the aptamer and distal AG has a strong influence on the splicing pattern. The mutations of pyrimidines upstream the stem in A3SS-T9 decreased the splicing efficiency. Interestingly, the presence of TetR still increases the A3SS recognition. Increased Py-tract downstream the aptamer leads to the recognition by spliceosome only of the distal 3SS and the A3SS is sequestered in the stem. However, the aptamer distance relative to the 3SS seems to be important for

switching properties. Recognition of A3SS in the presence of TetR is diminished with increasing distance between the aptamer and distal AG, which is observed in the constructs A3SS-T11, A3SS-T12 and A3SS-T13. Based on these results, the A3SS-T11 was selected as the constructs with the most desirable properties. In this construct, in the absence of TetR spliceosome recognizes exclusively the distal AG and the A3SS is sequestered in the aptamer stem. However, the presence of TetR activates the A3SS recognition. Again the A3SS can be switched off by addition of dox.

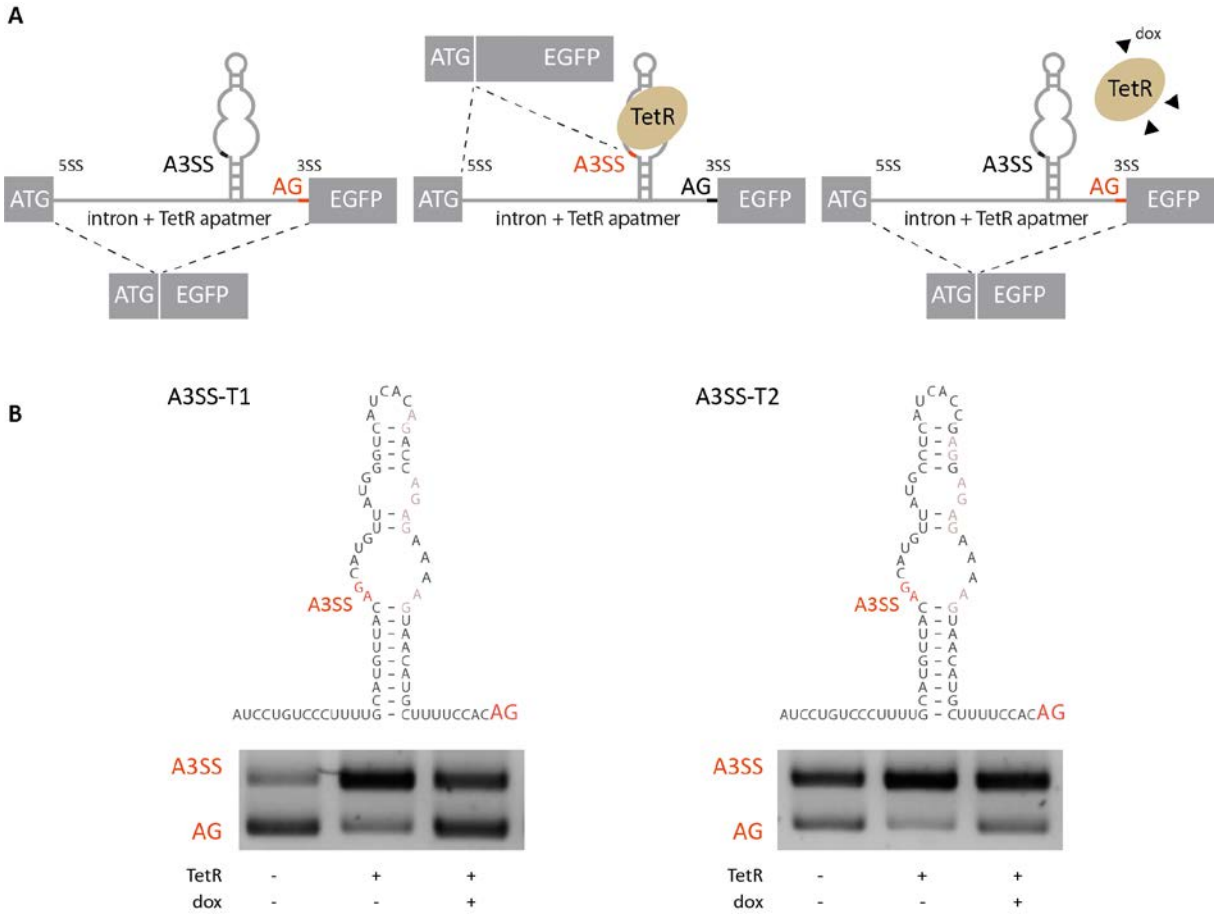


Figure 4.9 TetR aptamer placed at 3SS. (A) TetR aptamer with A3SS is placed upstream the canonical 3SS of MINX intron. (left) In absence of TetR, distal AG is recognized by spliceosome. (middle) In the presence of TetR, A3SS is recognized. (right) Splicing pattern can be restored by addition of dox that leads to conformational changes of TetR and the release of the aptamer. Selected 3SS is marked in red. Exons are displayed as boxes and intron with TetR aptamer as a line. **(B)** Schemas of the TetR aptamer structure at the 3' intron end of A3SS-T1 and A3SS-T2 construct. Used alternative A3SS and distal AG are marked in red. In the aptamer stem other alternative AGs are marked. Splicing pattern visualised by RT-PCR. HeLa cells were transiently transfected with the constructs A3SS-T1 and A3SS-T2 and co-transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. Upper bound corresponds to usage of A3SS and lower to distal AG. The spliced products were cloned using CloneJET PCR cloning kit (Thermo Scientific) and sequenced for verification. Experiment was repeated three times with similar results.

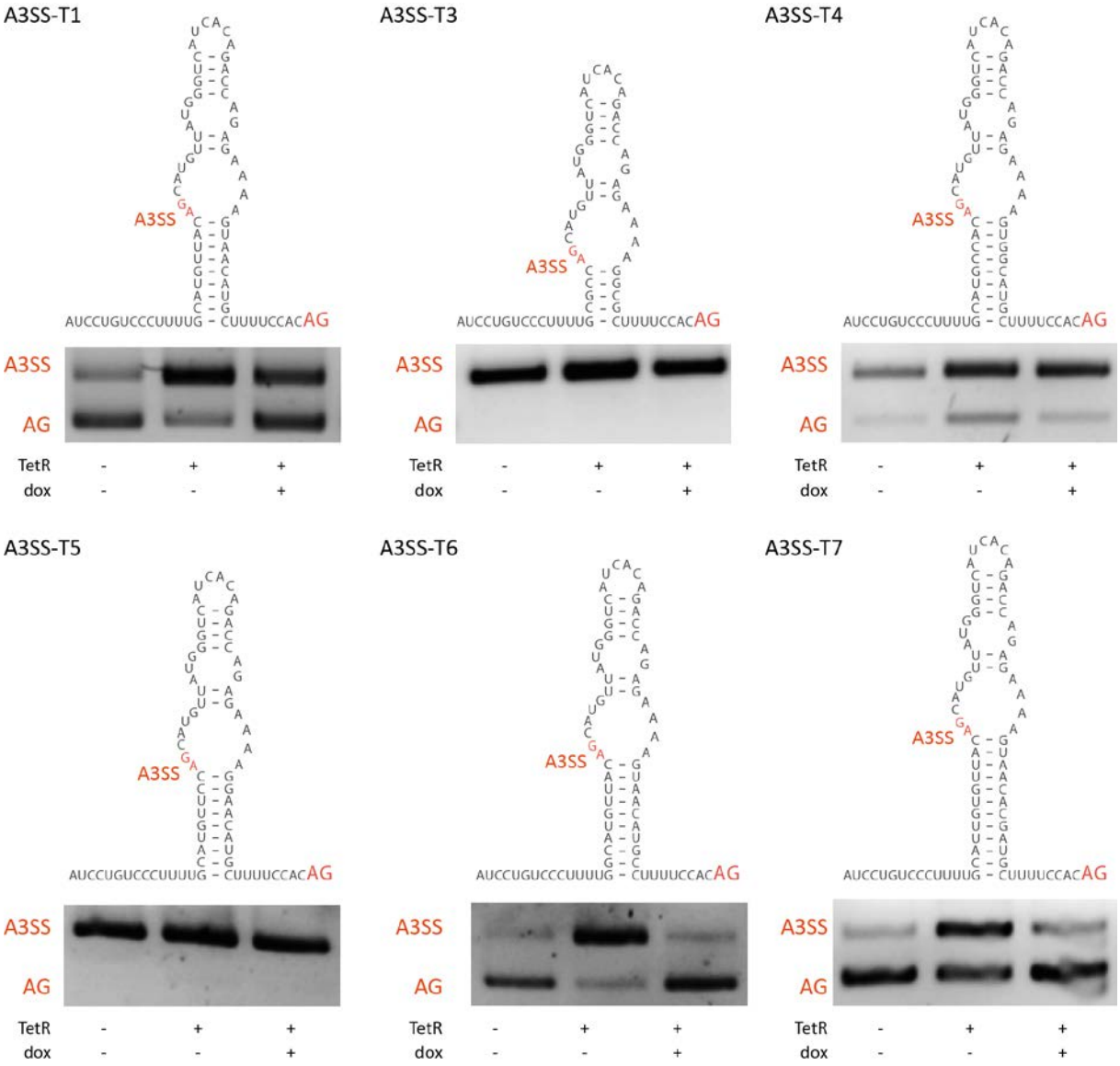


Figure 4.10 The aptamer stem modification. Schemas of the TetR aptamer structure at the 3' intron end of A3SS-T3-T7 constructs. Used alternative A3SS and distal AG are marked in red. Splicing pattern visualised by RT-PCR. HeLa cells were transiently transfected with these constructs and co-transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 μM dox for 24h. Upper bound corresponds to usage of A3SS and lower to distal AG. Experiment was repeated three times with similar results.

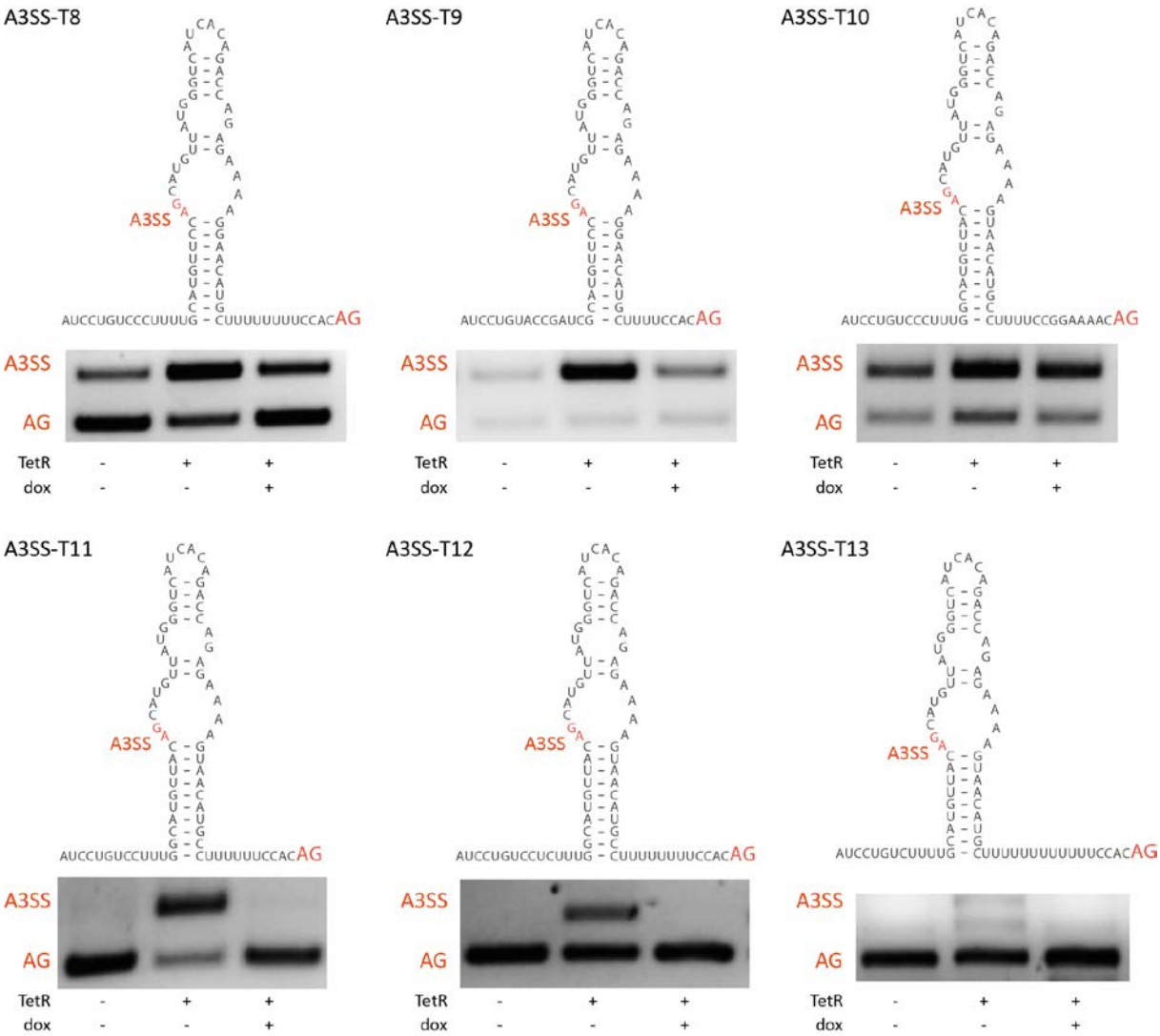


Figure 4.11 The Py-tract characterization. Schemas of the TetR aptamer structure at the 3' intron end of A3SS-T8-T13 constructs relative to the Py-tract position and composition. Used alternative A3SS and distal AG are marked in red. Splicing pattern visualised by RT-PCR. HeLa cells were transiently transfected with these constructs and co-transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Upper bound corresponds to usage of A3SS and lower to distal AG. Experiment was repeated three times with similar results.

Introduced mutations that destroys the TetR binding leads to a complete loss of A3SS regulation, indicating that switching is indeed mediated by TetR aptamer interaction (A3SS-T11*, Figure 4.12). Moreover, mutated A3SS in A3SS-T11 leads to exclusive usage of another A3SS* (A3SS-T14, Figure 4.12) in the aptamer and loss of switching behaviour. Additionally, lariat PCR to detect branch point (BP) sequence for both splice variants was performed. In absence of TetR the BP was predicted to be at adenosine (A) in the window of 42 nt upstream the distal AG. It is relative distance, where the stem length is not counted in to the final distance and the stem keeps 3SS closer to

the BP [128]. While, in the presence of TetR the BP was predicted to be at the same A in the window of 42 nt to A3SS.

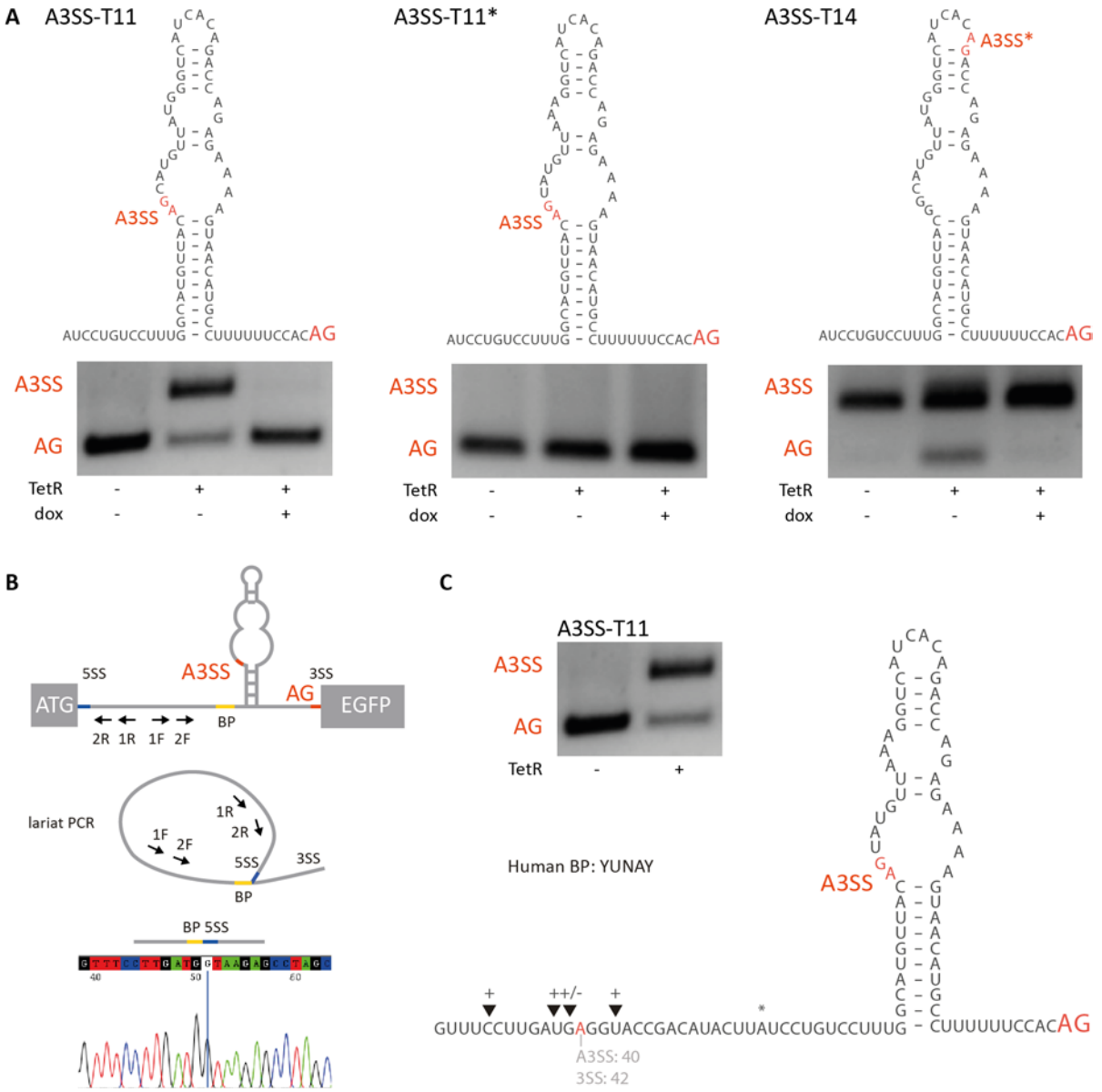


Figure 4.12 Mutated TetR binding site and the stem stabilization in A3SS-T11 and lariat PCR. (A) Schemas of the A3SS-T11* and A3SS-T14 constructs with mutated TetR binding site and the stabilized stem, based on A3SS-T11, respectively. Used alternative A3SS or A3SS* and distal AG are marked in red. Splicing pattern visualised by RT-PCR. HeLa cells were transiently transfected with these constructs and co-transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Upper bound corresponds to usage of A3SS and lower to distal AG. The spliced products were cloned using CloneJET PCR cloning kit and sequenced for verification. Experiment was repeated three times with similar results. **(B)** Schema of lariat PCR design for A3SS-T11 construct. Alternative A3SS and distal AG are marked in red, and 5SS and BP are marked in blue and yellow, respectively. The lariat PCR products were sequenced for verification and the sequencing chromatogram is depicted. **(C)** Splicing pattern visualised by RT-PCR and schema of detected BP positions. The triangles indicated detected putative BP in the presence of TetR (+) or in its absence (-). The predicted BP for MINX without the aptamer is marked as *. Human BP consensus is shown [129]. Additionally, the relative distance between predicted BP (A in red) and A3SS or distal AG is presented.

Following an existing alternative splicing mechanism for modulating the cell cycle-regulated Nek2 kinase localisation [130], that enables it to have both cytoplasmic and nuclear functions, we designed a device (A3SS-TN) for controlling nuclear import. Taking advantage of the control of alternative splicing in A3SS-T11 it is possible to produce two splice variants. The aptamer P2 stem can be tailored according to various required modifications while preserving its structure, specifically, the binding region. This allows for insertion of a nuclear localisation sequence (NLS) between the A3SS and AG. Which further would leads to production of different splice variants with different subcellular localisation on demand (Figure 4.13).

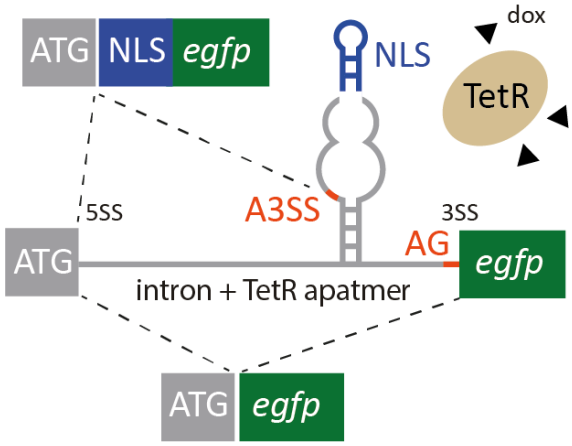


Figure 4.13 Schema of the proposed model to control nuclear import. In absence of TetR, distal AG is recognized by spliceosome. In the presence of TetR the A3SS is activated and furthermore, splicing pattern can be restored by addition of dox that leads to conformational changes of TetR and releasing the aptamer. Insertion of nuclear localisation sequence (NLS, in blue) between A3SS and AG allows to produce different splice variants with different subcellular localisation on demand. Exons are displayed as boxes and intron with TetR aptamer as a line.

First, it was assessed if the introduction of a NLS sequence between the A3SS and 3SS in A3SS-T11 would disrupt the switching properties (Figure 4.14). If consequently could lead to the production of different splice variants with and without NLS. To maintain NLS and EGFP in the open reading frame the stem length or distance between the aptamer and 3SS (A3SS-TN1, A3SS-TN2 and A3SS-TN3) was adjust. For these constructs slight splicing pattern changes were observed, however, A3SS-TN2 exhibits desired properties to produce two splice variants in the present of TetR. Interestingly, in these constructs also the ability to activated A3SS in the present of TetR was observed, due to the distance between the aptamer and distal AG.

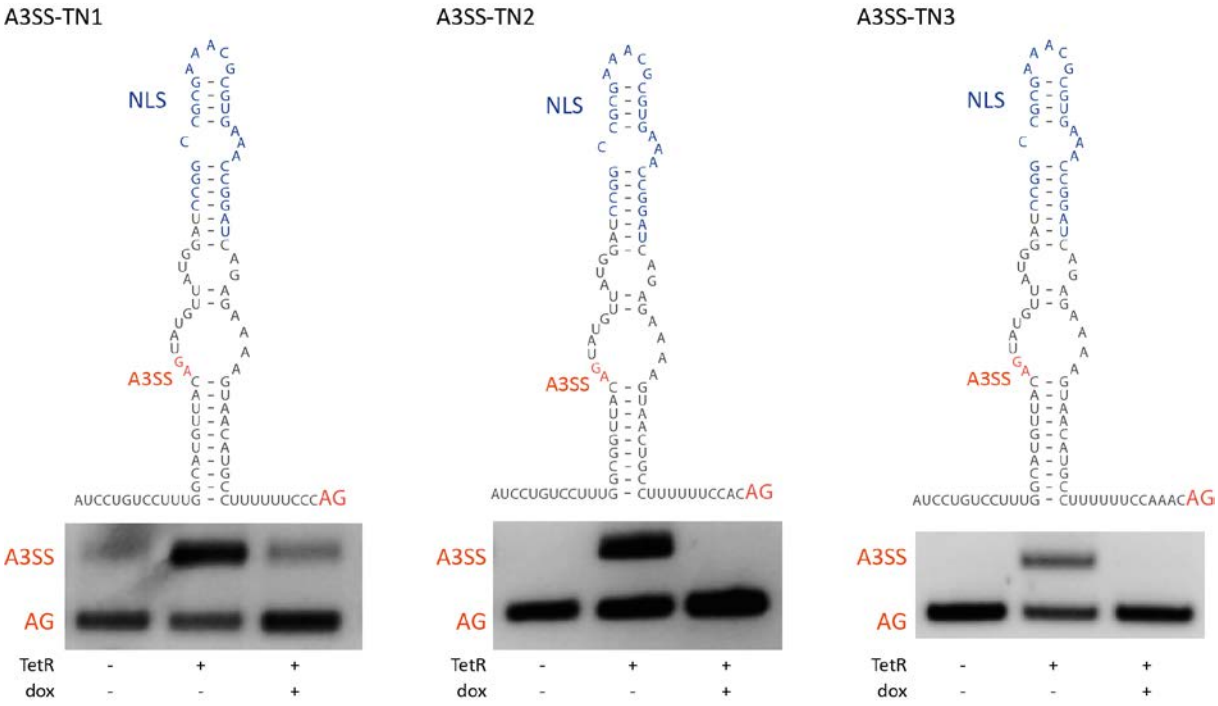


Figure 4.14 Insertion of NLS sequence within the TetR aptamer. Schemas of the A3SS-TN1 and A3SS-TN2 constructs with added NLS sequence (in blue) on top of the TetR aptamer. Alternative A3SS and distal AG are marked in red. Splicing pattern visualised by RT-PCR. HeLa cells were transiently transfected with the constructs A3SS-T1 and A3SS-T2 and co-transfected with plasmid expressing TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Total RNA was prepared and used for RT-PCR with primer pairs binding to both exons. Upper bound corresponds to usage of A3SS and lower to distal AG. The spliced products were cloned using CloneJET PCR cloning kit (Thermo Scientific) and sequenced for verification. Experiments were performed in triplicate and repeated three times with similar results.

Further, it was examined if the control on mRNA levels can be translated into protein level. For this purpose, A3SS-TN2 was fused with the cytosolic protein (CP) in order to prevent passive EGFP diffusion into the nucleus resulting in A3SS-TN2-CP construct. To reduce background of untransfected cells, A3SS-TN2-CP as well as EGPF control were stably integrated into a HeLa HF1-3 cell line using the Flp-In system. A generated cell line expressing the A3SS-TN2-CP was subsequently transiently transfected with TetR::mCherry construct that can be visualized by fluorescence microscopy. After transfection with a plasmid expressing TetR and treated with or without 50 µM dox for 24 or 48 h, cells where fixed and stained with DAPI. The microscopic images of EGFP show a homogenous distribution throughout the cell, however with a tendency to accumulate in the nucleus due to passive diffusion, while expressed TetR tagged with mCherry was primarily localized in the nucleus (Figure 4.15).

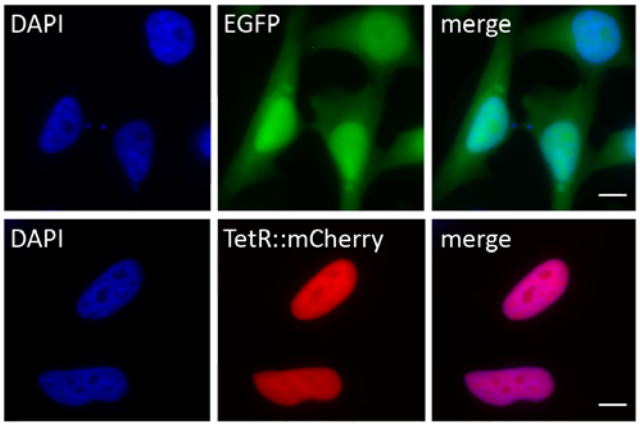


Figure 4.15 Microscopic visualisation of EGFP and TetR::mCherry controls. EGFP is stably integrated into HF1-3 cell line using the Flp-In system and transiently transfected Hela cells with the mCherry-tagged TetR. Cells where fixed and stained with DAPI. Scale bar 10 µm.

Constitutively expressed A3SS-TN2-CP show predominant cytoplasm localization (Figure 4.16A). It confirms that only distal the 3SS is used, producing an isoform without NLS. In the present of TetR, in most cells EGFP is expressed in the cytosol as well as it is accumulated in the nucleus and nucleoli. It indicates that the A3SS is activated in the presents of TetR and the second splice variant carrying NLS is produced (Figure 4.16B). Additionally, cells expressing TetR and incubated with dox show mostly cytoplasm distribution, meaning that dox leads to the release of TetR from the pre-mRNA (Figure 4.16C).

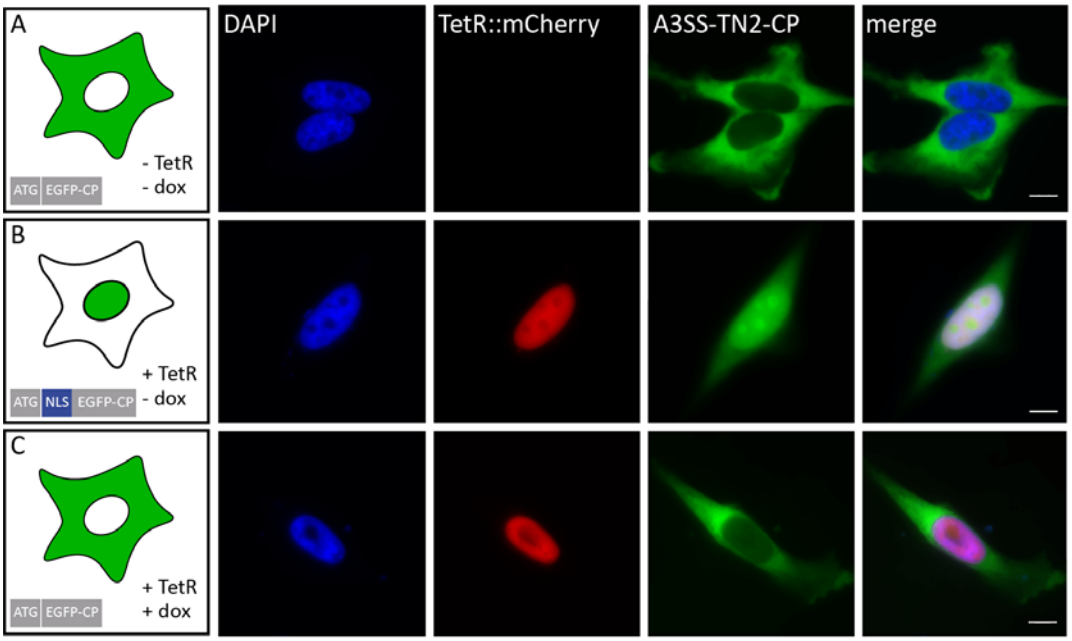


Figure 4.16 Microscopic visualisation of inducible control of nuclear import with TetR aptamer. (A) A3SS-TN2-CP stably integrated into HeLa HF1-3 cell line using the Flp-In system. (B-C) A generated cell line expressing A3SS-TN2-CP construct was transiently transfected with mCherry-tagged TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Cells where fixed and stained with DAPI. Scale bar 10 µm.

In sum, with the TetR aptamer complex it is possible to switch on A3SS on demand and moreover produce two functional splice variants with different cytosol and nuclear distribution. After 48 h most cells expressing A3SS-TN-2-CP and TetR show homogenous distribution of EGFP in the nucleus (Figure 4.17).

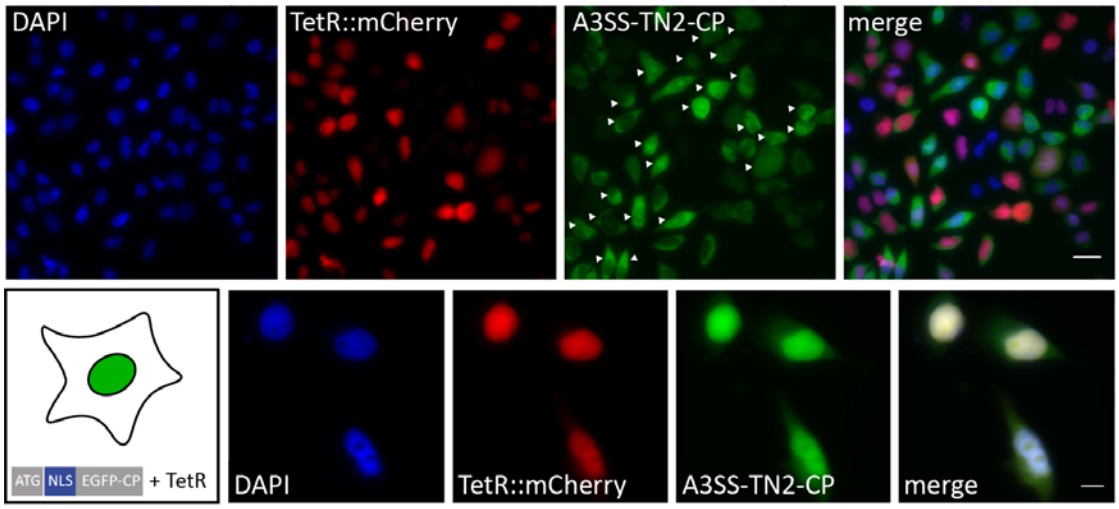


Figure 4.17 Microscopic visualisation of inducible control of nuclear import with TetR aptamer. A generated cell line expressing the A3SS-TN2-CP was transiently transfected with mCherry-tagged TetR construct and incubated 24h (upper panel) or 48h (lower panel). White triangles indicate the cells with observed EGFP accumulation in the nucleoli. Cells where fixed and stained with DAPI. Scale bar 25 µm (upper panel) and 10 µm (lower panel).

Additionally, it was assessed whether it is possible to produce two splice variants with mitochondrial or nuclear distribution. For this purpose, it was verified if a EGFP with a strong mitochondrial localization signal (MLS) from Cytochrome C Oxidase Subunit 8A (COX8A) will localize in mitochondria and if EGFP with both targeting signals, NLS and MLS is able to accumulate in the nucleus. Control expressing MLS-EGFP and NLS-MLS-EGFP were transiently transfected in HeLa cells. The microscopic images of the MLS-EGFP show a mitochondrial distribution and NLS_MLS_EGFP was primarily localized in the nucleus (Figure 4.18).

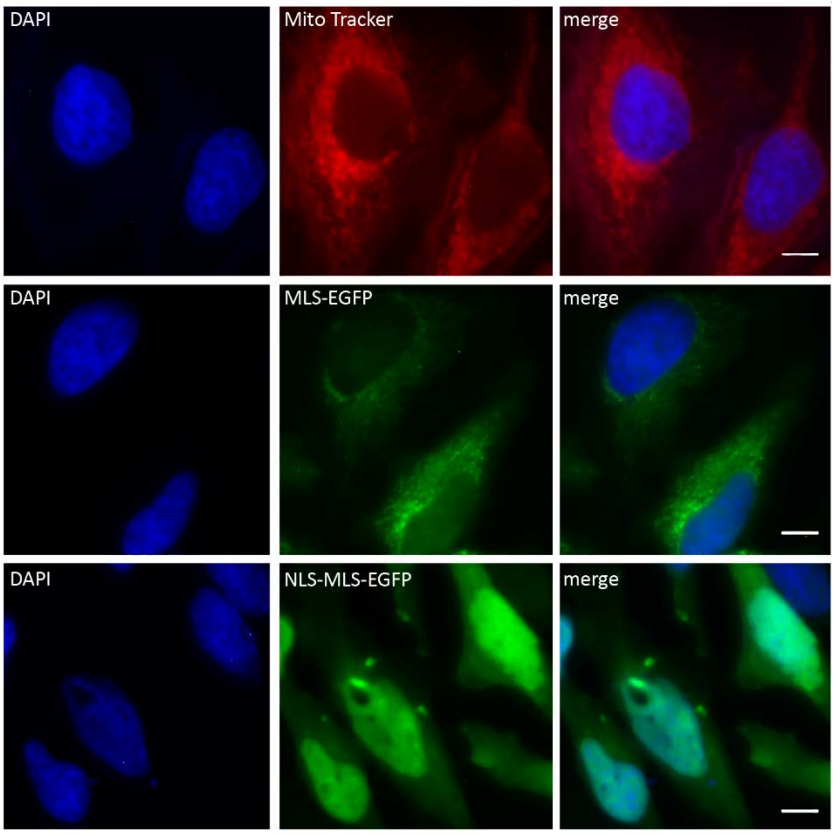


Figure 4.18 Microscopic visualisation of MLS-EGFP and NLS-MLS-EGFP controls. (Upper panel) Mitochondrial staining reagent was used to label mitochondria of HeLa cells (Mito Tracker, Abcam). Transiently transfected Hela cells with MLS-EGFP (middle panel) and NLS-MLS-EGFP (lower panel). Cells where fixed and stained with DAPI. Scale bar 10 µm.

Further, A3SS-TN2 was fused to the MLS resulting in the A3SS-TN2-MLS construct, that was stably integrated into HeLa HF1-3 cell line using the Flp-In system, to reduce background of untransfected cells. A generated cell line expressing the A3SS-TN2-MLS construct was subsequently transiently transfected with the mCherry-tagged TetR. After transfection with a plasmid expressing TetR and treated with or without 50 µM dox for 24 h cells where fixed and stained with DAPI (Figure 4.19). Constitutively expressed A3SS-TN2-MLS show predominant mitochondrial localization (Figure 4.19A). It confirms that only the distal 3SS is used, producing an isoform without NLS. However, in the present of TetR only mitochondrial EGFP distribution is observed and the nuclear localization is not detected. The same EGFP pattern is observed in the cells expressing TetR and incubated with dox (Figure 4.19B-C).

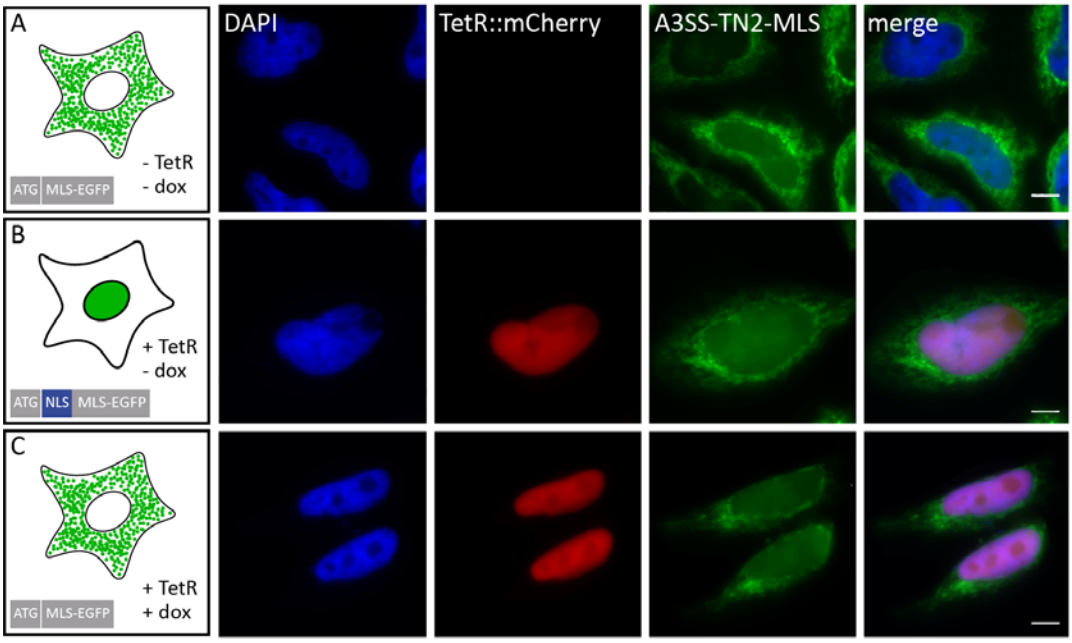


Figure 4.19 Microscopic visualisation of inducible control of nuclear import with TetR aptamer. (A) A3SS-TN2-CP stably integrated into HeLa HF1-3 cell line using the Flp-In system. (B-C) Generated cell line expressing A3SS-TN2-CP construct was transiently transfected with mCherry-tagged TetR (+) and treated with (+) or without (-) 50 µM dox for 24h. Cells where fixed and stained with DAPI. Scale bar 10 µm.

In a cooperation with the Cancer Metabolism research group, headed by Dr. Dimitrios Anastasiou at the Francis Crick in London we will test the model to investigate the function of *GLS* isoforms. This gene encodes the K-type mitochondrial glutaminase. Glutamine is an essential nutrient for cancer cell proliferation, especially in the context of citric acid cycle anaplerosis. Recent studies have shown the importance of glutaminolysis in maintaining the malignant phenotype, providing indications that glutaminases may be a potential therapeutic targets in some cancer cells [131]. The idea is to control the formation of the two splice variants of GLS with cytosolic and mitochondrial localization to assess the function of both isoforms in the glycolytic pathway. For this purpose, the aptamer stem will be replaced by the native MLS form GLS. While, instead of *egfp* is placed *GLS*. First, the predicted MLS sequence from *GLS* was tested if it has mitochondrial targeting properties. The fusion of the MLS (Table 6.1) to EGFP shows predominant mitochondrial localization in HeLa cells (Figure 4.20). Next, we designed and cloned constructs with MLS placed within the aptamer, like in the case of A3SS-TN2. These constructs were stably integrated and are currently tested in the Anastasiou's research group.

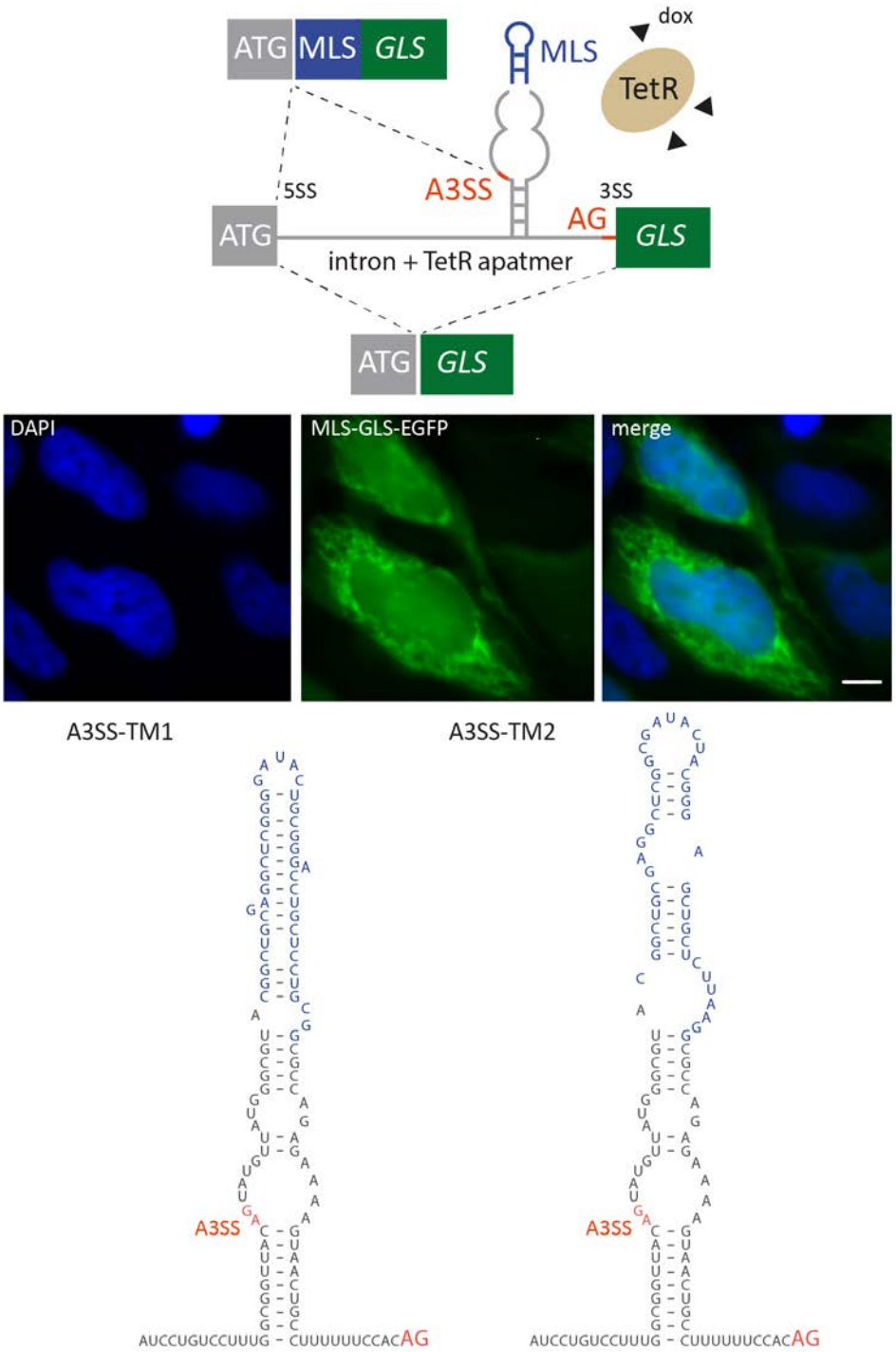


Figure 4.20 Control of GLS splice variants. (upper panel) Schema of the proposed model to control mitochondrial import. In absence of TetR, distal AG is recognized by spliceosome. In the presence of TetR the A3SS is activated and furthermore, splicing pattern can be restored by addition of dox that leads to conformational changes of TetR and the release of the aptamer. Insertion of mitochondrial localisation sequence (MLS, in blue) between A3SS and AG allows to produce different splice variants of GLS with different subcellular localisation on demand. Exons are displayed as boxes and intron with TetR aptamer as a line. (middle panel) Microscopic visualisation of transiently transfected HeLa cells with MLS-GLS-EGFP construct carrying the predicted MLS form *GLS*. Cells where fixed and stained with DAPI. Scale bar 10 µm. (lower panel) Schemas of the A3SS-TM1 and A3SS-TM2 constructs with added MLS sequence (nucleotides in blue) on the top of TetR aptamer. Alternative A3SS and distal AG are marked in red.

4.3 Project III: Control of mammalian translation with TetR aptamer

Translational regulation constitutes an important point of post-transcriptional control of gene expression, enabling the cell to change the level of gene product rapidly. Translation is mainly regulated at the step of initiation [132]. In eukaryotes, the 43S preinitiation complex is recruited at the 5' cap and scans the entire 5' untranslated region (UTR) in search for a start codon. The presence of secondary structure elements such as stem-loops or hairpins in the 5' UTR negatively affects translation initiation [133]. The potency of inhibition is dependent on the stability and the position of the structured elements in the 5' UTR. Insertion of aptamers downstream of the 5' cap and in front of the start codon can block ribosome scanning and translation of the downstream gene [134, 135, 136]. However, translational control differs in yeast and mammalian cells and this mechanism appears to only work in yeast, while the mammalian ribosome is able to efficiently scan through structured RNAs before the start codon [19, 137]. In this project, TetR aptamer dependent translational control system was developed and optimized in mammalian cells.

4.3.1 Results

The TetR aptamer was placed within the 5' UTR of the *egfp* reporter gene in a way that TetR binding to the aptamer efficiently inhibits ribosomal scanning. The translation start codon AUG is localized in the aptamer (Figure 4.21). Additionally, a 2A 'self-cleaving' peptide from the *porcine teschovirus-1* 2A (P2A) was attached upstream of the *egfp* to remove additional aptamer-born amino acids. P2A peptide results into one additional amino acid added to the protein of interest and it is expected that the majority of proteins may tolerate this minimal change of their protein sequence. TetR binding to the aptamer stabilizes its secondary structure which leads to inhibition of the ribosome scanning and consequently the inhibition of translation (OFF state) of the controlled gene. Addition of dox then releases TetR from the aptamer resulting in correct translation (ON state; Figure 4.23).

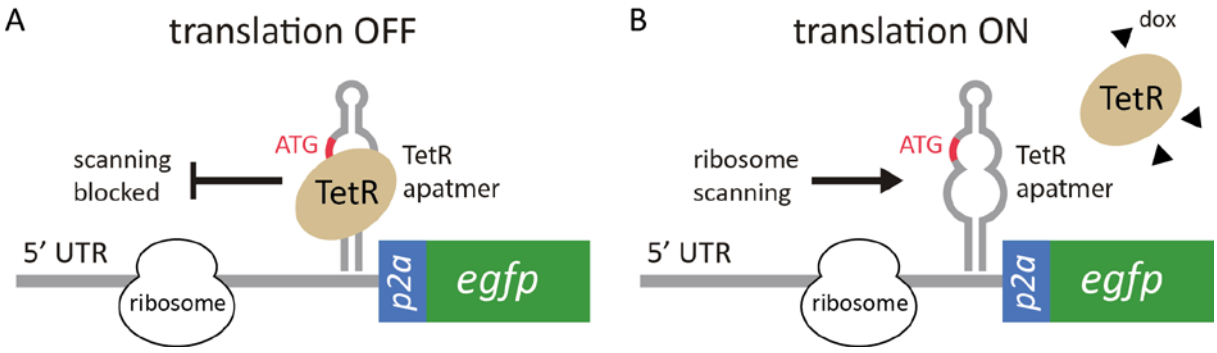


Figure 4.21 Translation control with TetR aptamer. TetR aptamer within an 5' untranslated region (UTR) of the *egfp* reporter gene. The start codon is marked in red and self-cleaving domain (*p2a*) in blue **(A)** In the presence of TetR, ribosome scanning is blocked and consequently no translation of the controlled gene takes place. **(B)** Translation can be restored by addition of dox that leads to conformational changes of TetR and the release of the aptamer.

To assess whether the TetR aptamer is able to control translation in HeLa cells, constructs T1, T2 and T3 (Figure 4.22A) were designed. The aptamer sequence contains two start codons, therefore one of them was mutated (T1 and T2). In addition the aptamer stem was stabilized by CG base pairs in T2 construct.

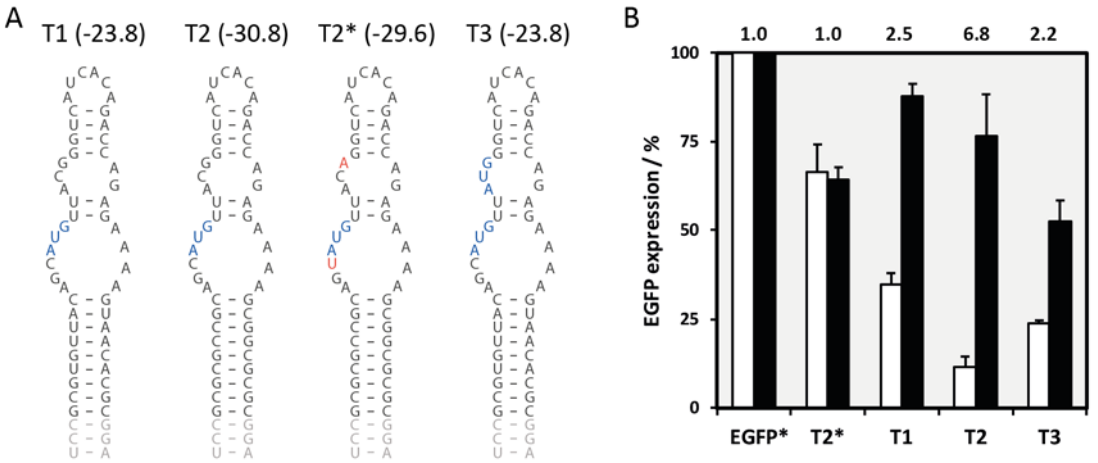


Figure 4.22 Assessing the control of translation with TetR aptamer. (A) Schemes of the T1-T3 constructs. The start codon is marked in blue and mutated nucleotides in red. A possible additional binding of 3 base pairs (in grey) of the BamHI restriction site with the P2A sequence was predicted with mfold program. Additional stem formation was included in the final ΔG prediction. The ΔG in kcal/mol of the constructs is displayed in brackets **(B)** Displayed are the EGFP fluorescence for all constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. Constructs were co-transfected with TetR:mCherry. EGFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry and mCherry expression was used to normalize for variation in transfection efficiency. EGFP* is a control without the aptamer. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

Cells were transiently transfected with these constructs together with a plasmid expressing TetR tagged with mCherry. EGFP and mCherry expression were monitored 24 h after transfection in the absence or presence of dox using flow cytometry. mCherry expression was used to normalize for variation in transfection efficiency. Construct T2 exhibits the best regulatory properties with a 6.8-fold (Figure 4.22B) and was chosen for further optimizations. Additionally, the introduction of the two mutations that destroy TetR binding (T2*) leads to a complete loss of regulation indicating that switching is indeed mediated by TetR aptamer interaction.

Furthermore, it was assessed whether and to what extent the stability of the aptamer influences regulation. Based on T2, a set of constructs T4-T31 was designed, mainly modifying the aptamer stem (Figure 4.23A) and the hairpin loop (Figure 4.23B). The aptamer stem was destabilized by stem truncations, exchanging a CG for weaker AU or GA base pairs, the insertion of an adenine bulges or an AG mismatch. While, insertion of additional CG base pairs lead to stem stabilization, the hairpin loop was exchanged by more stable tetraloops. The aptamer structures were designed with predicted stabilities between -14 and -42 kcal/mol. The ΔG (kcal/mol) was calculated for the whole aptamer sequence using the mfold program.

Only T6 and T19 of all tested constructs show improved regulation (8.5-fold). Interestingly, the removal of the three base pairs in the closing stem of the aptamer results in an increased *egfp* expression in the presence of dox with a maintained low basal expression level in the repressed state. All constructs with the destabilized aptamer stem exhibit higher *egfp* expression with or without dox, in contrast to the construct with the stabilized aptamer stem. An too weak the aptamer stem is associated with the loss of translational control by TetR. Moreover, the stabilization of hairpin loops leads to decreased *egfp* expression in both states but neither of them showed increased regulation (Figure 4.23B).

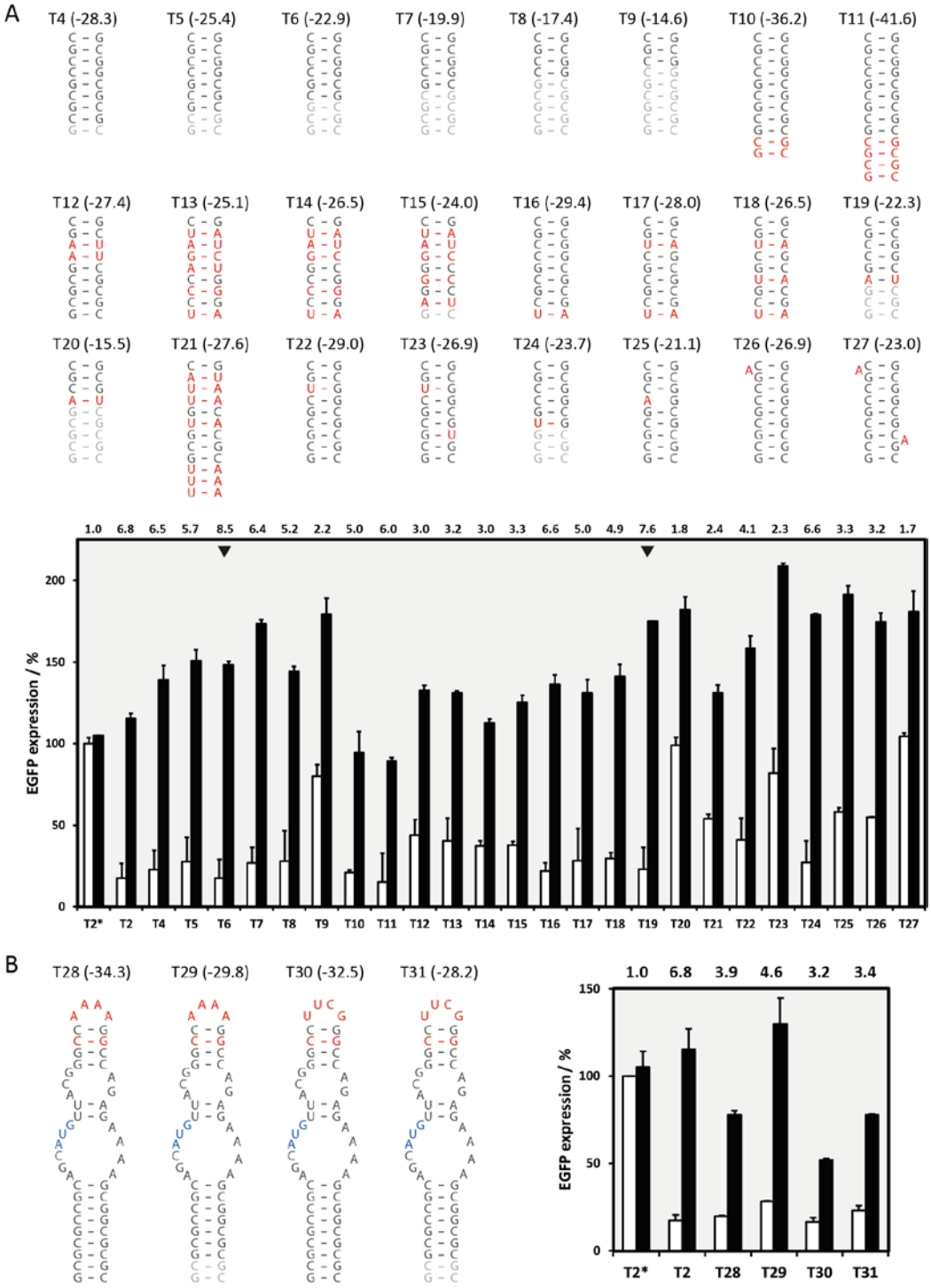


Figure 4.23 T2 optimization. (A) Schemes of the T4-T27 constructs with modified aptamer stem and corresponding *egfp* expressions. Deleted nucleotides are marked in light grey and in red introduced mutations. The triangles indicate constructs with improved fold (T6 and T19). **(B)** Schemes of the T28-T31 constructs with modified aptamer closing loop and corresponding *egfp* expressions. The start codon is marked in blue. (A-B) Values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. The T2* expression was set as 100%. Constructs were co-transfected with TetR:mCherry. EGFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry. mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

It is known that the extent of inhibition is a function of the stability and position of the structured element in the 5' UTR [133]. To study the position and structure dependence of the aptamer-mediated regulation, the aptamer was placed directly behind the cap site (T32; Figure 4.24A). The cap-proximal aptamer structure is highly inhibitory (Figure 4.24D), even in the absence of TetR. Although, the translation regulation is observed with the aptamer but with no improved regulation.

Next, was assessed whether the presence of short uORF (upstream open reading frame) can influence and increase the regulation of translation of the main ORF with the aptamer. Construct T33 contains a stop codon downstream the aptamer, and a second start codon placed upstream of the *egfp*, generating an uORF (Figure 4.24B). However, the presence of the uORF in this construct has only a modest effect on the translation regulation of main ORF (Figure 4.24D).

In yeast, it has been demonstrated [138] that a ligand inhibition of mRNA translation is most efficient when multiple binding sites are present in the 5' UTR. Therefore, effect of tandem aptamer on the dynamic range in HeLa cells was studied. TT1-TT3 constructs were designed with two and TT4 with three aptamers within 5' UTR (Figure 4.24C). Additionally, in TT2 a second start codon in the aptamer was shifted out of the ORF. TT1 and TT2 have the cap-distal aptamers positioned, while TT3 contains the aptamers at both 5' UTR ends. In the case of TT4 construct the aptamers are cap-proximal, middle and distal located. The use of two aptamers in TT1 and TT2 slightly reduced the *egfp* expression as well as the regulation (Figure 4.24D). However, the cap-proximal and -distal aptamer positions in TT3 led to an increased *egfp* expression in the presence of dox with maintained low basal expression level in the repressed state, consequently improving the regulation. Likewise, insertion of three aptamers within the 5' UTR in TT4 shows improved regulation (11.2-fold).

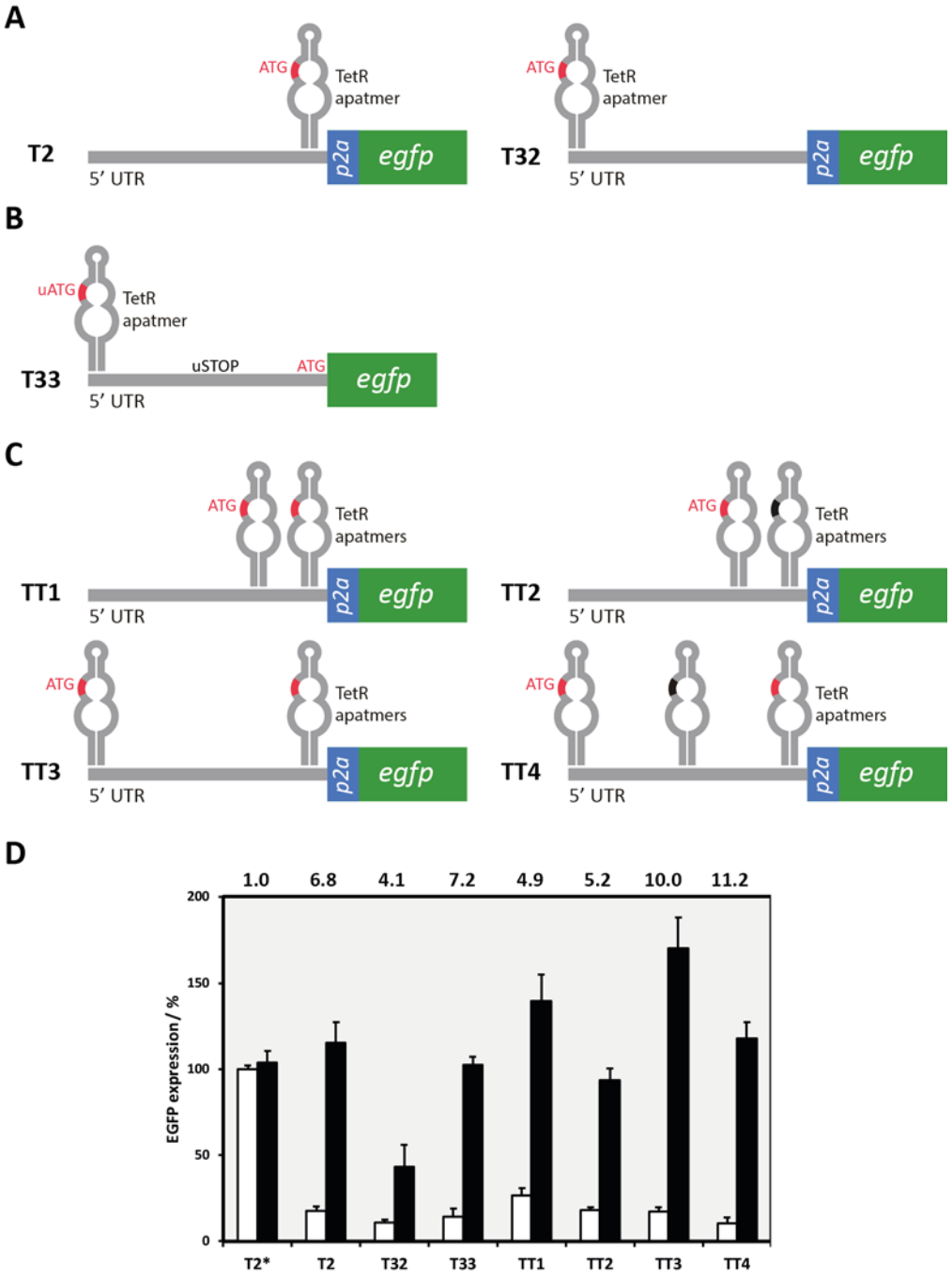


Figure 4.24 Attempts to improve a translation regulation with the aptamer. Constructs with the **(A)** cap-distal or proximal aptamers location, **(B)** introduced upstream open reading frame (uORF) and **(C)** multiple TetR binding sites in the 5' UTR of the *egfp* reporter gene. The start codon in the ORF is marked in red. **(D)** The EGFP fluorescence for all constructs, values without (white bars) and with dox (black bars) and above each construct the regulatory activity are shown. The T2* expression was set as 100%. Constructs were co-transfected with TetR:mCherry. EGFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry. mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

Additionally, this system was tested together with the TetR aptamer splicing device from the first project to achieve a tighter control of gene expression. The aptamer from T2 construct was inserted into the 5' UTR of SP*, generating T2-SP* constructs (Figure 4.25). Dual control of *egfp* expression at the splicing and translation level provides increased regulation up to 14.7-fold of controlled gene, compared to an individual control (~7-8.5-fold).

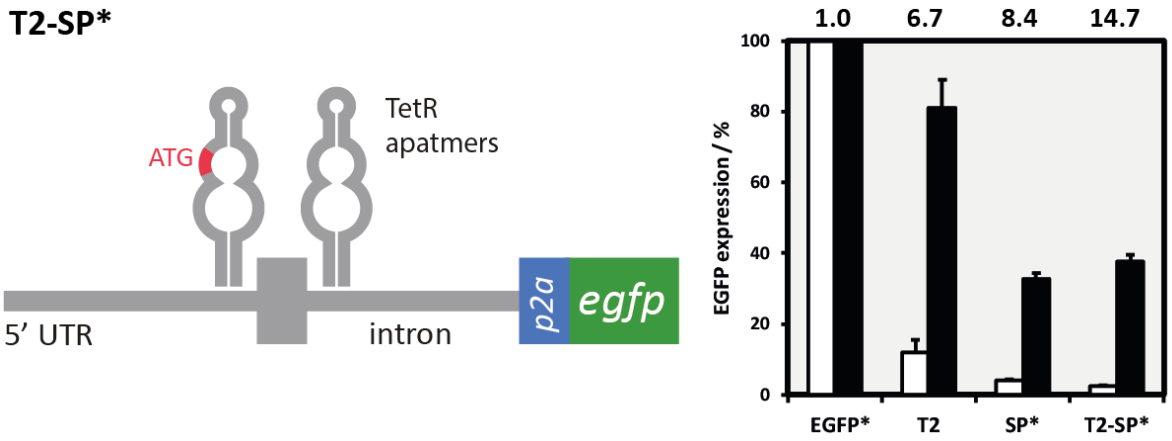


Figure 4.28 Dual control of a gene expression. (left) Schema of the T2-SP* construct with TetR aptamer inserted within the 5' UTR and intron. Exons are displayed as boxes, 5' UTR and intron with TetR aptamer as a line. The start codon in the ORF is marked in red. (right) The EGFP fluorescence for EGFP*, T2, SP* and T2-SP* constructs, values without (white bars) and with dox (black bars) and above each construct are shown the regulatory activity. The EGFP* expression was set as 100%. Constructs were co-transfected with TetR:mCherry. EGFP and mCherry expression was measured 24 h after transfection in the absence or presence of 50 µM dox with flow cytometry. mCherry expression was used to normalize for variation in transfection efficiency. Experiments were performed in triplicate and repeated three times with similar results. Error bars represent the standard deviation from the means.

5 Discussion

5.1 Universal and efficient splicing device for controlling a gene expression in human cells.

In the first project, a versatile and highly efficient TSD was designed to control gene expression with the TetR aptamer in human cells. It has been reported that the 5SS is recognized by base pairing to the 5' end of U1. Previously, in our group inhibition of 5SS by artificial tc aptamer has been shown in yeast [60]. The tc aptamer is inserted in a way that the consensus sequence of 5SS was integrated within the closing stem of the aptamer and the 5SS in the presence of tc is blocked. Similar to this model, the TetR aptamer complex can conditionally block the U1 binding, which leads to intron retention. Reversible and robust control of 5SS recognition using TetR aptamer allowed us to establish the first intron retention switching module, named TSD, to control protein expression on the level of splicing. In addition, induction of TSD is achieved using inexpensive, cell permeable and not-toxic dox. The engineered switch shows a high dynamic range up to 10-fold and low basal activity that is a main challenge in mammalian synthetic biology. Additionally, TSD is an easily applicable device as it has only 292 nt, and a controlled protein of interest has only one added amino acid at the N-terminus. These characteristics represent a considerable improvement compared to existing RNA splicing regulators in human cells, which exhibits a low dynamic range up to 4-fold, are the size of ~1300 nt and are not reversible [61]. In addition, TSD was functional in HEK293, HeLa, A549 and CHO cell lines, thus demonstrating its robustness, despite the fact that splicing is highly cell type-specific.

Modularity is an essential concept in engineering fields that can be applied to synthetic biology. However, engineered devices may not actually exhibit modular behaviour and often a device's features may change under different conditions [106, 139, 140]. Designing the TSD, an influence of surrounding genetic context of the aptamer on its switching behaviour was observed. First, a cryptic 5SS in the aptamer closing stem was responsible for low expression of both FLuc and GFP genes. RT-PCR analysis confirmed significant accumulation of pre-mRNA and low mRNA level for these constructs. Moreover, inhibition of intron removal was caused also due to the binding of strong exonic splicing silencers directly upstream the 5SS. Mutations in both cryptic 5SS and predicted ESS motif significantly exhibit higher GFP expression. However, loss of the ESS motif is associated with the weakened regulation properties. Besides

this, a predicted small stem loop structure upstream the aptamer may contribute to the high switching regulation in SP* and additionally confirm the assumptions about the impact of splicing factors on the recognition of 5SS as often *cis*-elements are exposed in the loop stems [141]. These studies allow for better understanding and will guide future design of artificial splicing systems for controlling gene expression.
In sum, the proper functioning of the switching module was assessed in the context of different reporter genes like FLuc, GPF, and mCherry and the human transcription factor MAX.

The system may also be applied to gain insight into diverse cellular mechanisms, such as pre-mRNA splicing, to study the impacts on cellular behaviour. We anticipated that the TSD can be used in a multi-gene context, where several genes are controlled by TSD and these genes are spaced by additional 2A sequences. Additionally, TSD can be easily combined with other regulatory systems to control gene expression at different levels to achieve tighter control of gene expression compared to either individual module, e.g. with a tetracycline aptamer controlled 3SS or aptazymes that impact mRNA stability.

5.2 Control of an A3SS recognition with the TetR aptamer

With the TetR aptamer complex it is possible to switch on an A3SS on demand and moreover produce two functional splice variants with a different cytosol and nuclear distribution. In the absence of TetR, the spliceosome recognized only the distal 3SS and the A3SS is sequestered in the stem in accordance with the model for 3SS recognition [128].
The 3SS is encoded only by the dinucleotide AG, which is present in the genome very frequently. In mammalian genes the mechanism of 3SS selection is still poorly understood [142]. The spliceosome is assisted in the 3SS identification by a Py-tract, a sequence located between the branch site and the dinucleotide AG [143]. Most studies on the recognition of the 3SS site was done for yeast and are only little information for human cells. This is connected tied to the flexibility of 3SS, BP and Py-tract in mammalian introns. In yeast the BP sequence is highly conserved and early steps of splicing are independent of the 3SS. In mammalian splicing, independence of AG is related with the BP and Py-tract elements. The second step of splicing involves a mechanism in which the spliceosome scans the intron from the BP until the first AG is

encountered [91]. Mutations in the dinucleotide AG in mammalian genes prevent cleavage at the 5SS and spliceosome complex formation. This suggests that the 3SS site selection likely occurs in the first step of splicing [144]. Work of the Vilardell's group documented the rules that govern 3SS selection in *Saccharomyces cerevisiae*, revealing a widespread role for the RNA secondary structure in the intronic region between the BP and the 3SS in this selection (Figure 5.1A). In yeast, a 3SS has to be located within a window of 10 to 45 nucleotides from the BP, either linearly or via folded RNA. All splice sites within this window, but outside a stem, will be used by the spliceosome machinery with a preference for YAG. A region of 9 nt downstream the BP will not harbour any structure and likewise AGs located here will be ignored [128].

In A3SS-T11, the 3SS and A3SS are located in the window of 45 nt, based on the detected BP (Figure 4.12C). In human genes, the Py-tract plays an essential role in the splicing process and is mostly absent in yeast. It acts at an early stage of spliceosome assembly point and it is required for the first step of the catalytic process. The length of the Py-tract and its composition are important in 3SS recognition [145, 146]. It was observed that Py-tract location relative to the aptamer position as well as their nucleotide composition highly influences the selection of the 3SS and significantly affects the switching properties (Figure 4.11).

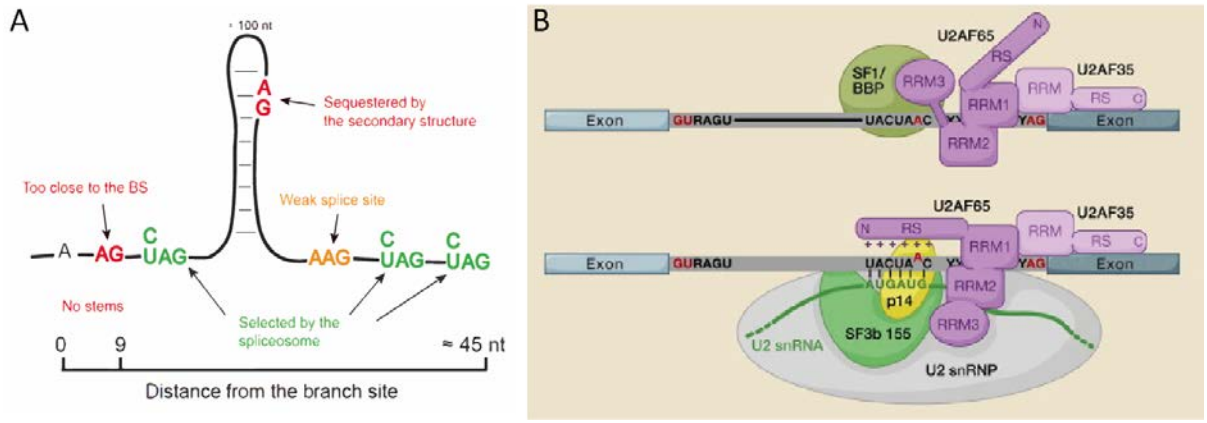


Figure 5.1 Mechanisms of 3SS recognition. (A) 3SS definition in *Saccharomyces cerevisiae* introns. Numbers indicate the distance to the branch point sequence (BP) A in nucleotides (nt). Stems are not considered by the spliceosome when determining this distance, and included AGs (red) are therefore not used. AGs preceded by a pyrimidine (green) are favoured over AAG (orange), otherwise the latter are used efficiently. Predicted stems in the first 9 nucleotides after the BP are not compatible with splicing, and AGs in this zone are not used as 3SS (adapted from [128]). **(B)** Molecular interactions at the 3SS within the spliceosomal E (upper panel) and A (below panel) complexes (adapted from [90]).

For A3SS-T11, the Py-tract is located between the aptamer stem and 3SS, where the stem is likely required to reduce the distance between the BP and the 3SS. Also Py-tract can be separated by the aptamer in A3SS-T1, where the aptamer presumably brings together Py-tract sequences to facilitate the recognition to binding proteins to these sequences such as the U2AF65 (Figure 5.1B). Unexpectedly, though, upon TetR binding to the aptamer, A3SS is activated. It contradicts other works where it was shown that the recognition of 3SS is blocked when ligands bind to the aptamer [58,61]. Presumably the aptamer position at the 3' end of the intron interferes with the SF1-U2AF65 and U2AF65-U2AF35 interaction during spliceosomal assembly. The spliceosome scanning mechanism from the selected BP may be blocked by the TetR-RNA complex resulting in a yet unexplained recognition of A3SS by the spliceosome machinery. Probably the accumulation of splicing factors next to the stem lead to its opening, as the same splicing pattern is observed for A3SS-T14 construct with a stabilized aptamer stem, where A3SS is recognized even without TetR, but it does not exhibit switching behaviour (Figure 4.12A). As in mammalian genes the mechanism of 3SS selection is still not fully understood, the application of aptamers may provide an additional strategy to study mechanism of alternative splicing in these cells.

5.2.1 Evaluation of the proposed model to control nuclear import

The TetR aptamer system to control an A3SS was established in the context of e*gfp* reporter gene. Following a successful demonstration of the producing of two splice variants on the mRNA level, this model was also confirmed on the protein level. Proteins smaller than 60 kDa or up to a diameter of 9 nm can passively diffuse into the nucleus [147]. EGFP is a protein with a molecular mass of 27 kDa and diameter of 2.4 nm [148]. Therefore, the strategy was to fused EGFP to a cytosolic protein in order to prevent passive diffusion into the nucleus. The EGFP fused to the SERPINE1 mRNA-binding protein 1 (SERBP1) with a molecular mass of 122 kDa shows predominant cytoplasm localization (Figure 4.16). This protein contains arginine and glycine rich motifs, which may lead to the accumulation inside nucleoli when it is transported into the nucleus [149, 147].

The system could be tested with different reporter gene that does not migrate to the nucleus via passive diffusion. Additionally, the EGFP fusion to other cytosolic proteins such as GTPase activating protein (SH3 domain) binding protein 1 (G3BP1) or adenosine monophosphate deaminase 2 (AMPD2) could be tested. Also, it was

observed that the TetR concentration based on mCherry expression in the cells has an impact on the switching properties. In cells, with an overly low or overly high TetR expression, it was not detected EGFP accumulation in the nucleus. An insufficient TetR concentration does not lead to activation of A3SS and therefore does not produce a nuclear splice variant. In turn, overexpression of TetR can lead to metabolic stress and pleiotropic effects in the cells.

Unfortunately, the attempt to produce two splice variants with mitochondrial or nuclear distribution with TetR system proved unsuccessful. Based on the RT-PCR results for A3SS-TN2-MLS, the A3SS is activated in the presence of TetR and second splice variant carrying MLS is produced. The microscopic images of the NLS-MLS-EGFP control show EGFP primarily localized in the nucleus. It is possible that the additional aptamer-born amino acids lead to not proper folding and masks the NLS [150] or NLS-MLS sequence may be cleaved off upon import into mitochondria. Most of the N-termini of mitochondrial proteins carrying a MLS are cleaved off by the heterodimer mitochondrial processing peptidase in the matrix and some of them are further cleaved by intermediate peptidase such as Oct1 or Icp55 [151, 152]. Additionally, the strong COX8A could be exchanged by the weaker mitochondrial signal.

Finally, in theory any targeting signals involved in protein relocalization can be controlled by our device including signals for nuclear export, mitochondrial, peroxisome signal, or any sequence of interest. This model could also be used for functional characterization of individual proteins involved in different cellular functions, determining mislocalization of disease-causing mechanisms.

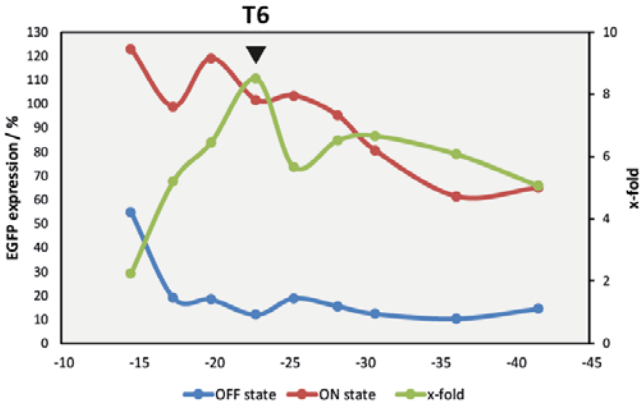
5.3 Blocking a ribosomal scanning with the TetR aptamer complex

Ribosomal scanning through structural barriers within the 5’ or 3’ UTRs of eukaryotic mRNA transcripts was found to be an important regulatory step [153, 154, 155]. RNA structures located in the 5’ UTRs are important to this regulation by affecting ribosomal recruitment and positioning at a favourable start codon. The main rate-limiting step of translation initiation is determined by the binding of the 43S pre-initiation complex (composed of the 40S ribosomal complex, initiation factors eIF3 eIF1, eIF1A, eIF5, and eIF2-GTP-met-tRNA) to mRNA via the eIF4 initiation factor complex (eIF4E, eIF4A, eIF4G, and eIF4B) [156, 157]. The 5’ methyl G cap is recognized by initiation factor eIF4E.

Under favourable translation conditions, eIF4G serves as a scaffold bridging the ribosome to the mRNA cap by binding eIF4E, eIF4A, and eIF3. Initiation factor eIF4A exhibits RNA helicase activity and is thought to assist the eIF4F complex in unwinding mRNA secondary structure, creating a binding site for the 43S initiation complex. The 43S complex then scans along the mRNA until it reaches a start codon.

Previous studies have shown that the introduction of a small molecule binding aptamer into the 5′ UTR of an mRNA can confer regulated expression of both prokaryotic and eukaryotic reporter genes [19, 50]. However, work on tetracycline and neomycin riboswitches indicated that the transfer of a yeast-optimized variant to higher eukaryotes is challenging, probably due to the increased helicase activity of the ribosome in mammalian compared to yeast cells. Translation control with the ciprofloxacin aptamer only exhibits up to 1.8-fold in HeLa cells [136], which indicate that the full potential of aptamer mediated translation control in human cells has been far from realised to date.

Results in this project are consistent with a model in which translational initiation is blocked by the aptamer-complex in the 5′ UTR. Placing TetR aptamer within 5' UTR of *egfp* reporter gene allows for reversible and efficient control of translation in HeLa cells. Interestingly, the aptamer structure in the 5' UTR in T6 construct did not reduce the basal expression (and only marginally in TT4) in the absence of TetR compared to the EGFP control without the aptamer.



Construct	ΔG	OFF state	ON state	x-fold
T10	-41.6	14.5	64.9	5.07
T11	-36.2	10.4	61.3	6.08
T2	30.8	12.4	80.5	6.67
T4	-28.3	15.6	95.2	6.52
T5	-25.4	18.9	103.4	5.67
T6	-22.9	12.1	101.6	8.50
T7	-19.9	18.4	119.0	6.45
T8	-17.4	19.2	98.9	5.20
T9	-14.6	54.9	122.9	2.25

Figure 5.2 Evaluation of translation regulation with the aptamer. Display of the constructs with different aptamer thermal stability (ΔG kcal/mol) and EGFP fluorescence in the presence of dox (ON state) or in its absence (OFF state) and dynamic range (x-fold). The triangle indicates T6 construct with the highest regulation.

In regard to aptamer thermal stability, it is was observed that increasing predicted thermal stability of mRNA leads to a decrease in translation (T2-T11; Figure 5.2), which agrees with previous studies [52, 158]. The steepest fall-off in translation efficiency occurs

when aptamer-predicted stabilities increase from -28 to -42 kcal/mol and the translation efficiencies remained consistent between the stabilities of -14 and -25 kcal/mol. The majority of constructs exhibit an EGFP fluorescent between 10-20% in the repressed state. Only T9 construct with the shortest aptamer stem shows approximately 55%. Moreover, most of the constructs exhibit regulation from 5- to 6.7-fold, and the maximum 8.5-fold is observed for T6 with -23 kcal/mol aptamer stability. Interestingly, the removal of three base pairs of the closing stem of T2 results in an increased dynamic range. This may suggest that the initial design of the aptamer stem with a length of 9 base pairs would show the highest switching properties if not the influence of possible additional binding of surrounding sequences. Mutations that destroy TetR binding were introduced in construct T6 with the best regulation properties (T6*). It leads to a complete loss of regulation indicating that switching is indeed mediated by TetR aptamer interaction. Additionally, for T6 and T6* constructs qPCR analysis were performed and no significant changes were observed in the present of TetR and with or without dox on mRNA level [data not shown]. This confirms that the control occurs at the translational level and is due to a specific interaction between TetR and the RNA aptamer.

In this study the significance of the aptamer position on translation was observed. Similar effects that in mammalian cells mRNA structure is more inhibitory when is proximal to the 5' mRNA cap were already shown in others studies [153] (T2 and T32, displayed in Figure 4.26). Reduced *egfp* expression in T32 exhibits lower regulation than T2, which can be related due to movement of both the inhibitory structure at the 5' cap proximity and the start codon, which is located within the aptamer sequence. Constructs TT1 and TT2 that contain two aptamers at the cap-distal position do not lead to tighter translation repression. The aptamers in TT1 and TT2 are separated by 30 nt spacer, that does not form any secondary structures and may thus be seen solely as an insulation module to minimize global folding constrains between the two RNA architectures [159]. However, it is still probable that an additional global- or self-folding between the same aptamer stems occurs, which could explain the lower regulation compared to the single aptamer. In contrast, the cap-proximal and -distal aptamers placement in TT2 as well as the additional third aptamer in the middle of the 5' UTR has higher effect on regulation control. It is possible that different steps of translation initiation are being inhibited by TetR, depending on whether the aptamer is cap-proximal or distal. Cap-proximal insertion may inhibit binding of the 43S complex to the

cap structure whereas the cap-distal aptamer may interfere with formation of the 80S ribosome, presumably by blocking the scanning preinitiation complex [33, 160].

In addition to thermal stability and distance effects, translation efficiency was dependent on the GC content of the aptamers closing stem. Constructs T1 and T6 both have predicted thermal stability around -23 kcal/mol but differ in GC content. Notwithstanding, T6 with high GC content shows regulation up 8.5-fold and T1 only regulation up 2.5-fold. The same pattern is observed for constructs T4 and T21 with predicted thermal stability of -28 kcal/mol, but their regulation is up to 6.5- and 1.7-fold, respectively. In has been reported that the GC content affects protein-translation efficiency independent of hairpin thermal stability and hairpin position [158]. It seems that local stability per base has an effect on translation efficiency, since compact, thermally stable hairpins composed of GC bonds are harder to melt than hairpins composed of weaker AU bonds. Presumably RNA structures with predicted thermal stabilities stronger than -50 kcal/mol may still be efficiently translated as long as the GC content of the stem is relatively low. Less stable RNA structures with high stem GC content may inhibit translation more than stable structures with low GC content. The scanning ribosome does not melt the entire hairpin in a concerted event, but rather tries to bulldoze through the obstacle in a progressive, relatively local unzippering. High GC content confers high stability per base pair, perhaps exceeding the force available to the translation machinery [158].

In this project an efficient translation control system in mammalian cells was developed that meets the desirable features that often prove a major challenge in implementing synthetic biology in mammalian cell lines. The TetR aptamer based translation regulation system could be broadly applicable for reconstructing and interrogating native translation control mechanisms, as well as for designing novel cellular functions. Modularity and applicability of the system have been demonstrated by regulation of endogenous genes as well as with the TetR aptamer in the context of other 5’ UTRs. Another approach would be to combine the TetR aptamer based translation regulation system with other systems that control gene expression at transcriptional or post-transcriptional level to achieve a tighter control of gene expression as was demonstrated by the dual control at the splicing and translation level. In summary, these studies allow for better understanding how translation is regulated in human cells and in addition could be useful in the design of artificial systems for controlling translation process in mammalian cells.

5.4 Transcriptional- versus post-transcriptional-based regulatory systems

Synthetic biology has pioneered transformative genetic devices that enable the study of cellular and molecular biology in cells. Mammalian regulatory devices use diverse mechanisms to allow flexible, precise, and comprehensive control over gene expression and cellular development [10]. Fine tuning of gene expression is critical for many synthetic biology applications and can be achieved at the transcriptional and post-transcriptional levels. Transcriptional control received the most attention in terms of synthetic devices to regulate gene expression in preference to other levels of control. Transcription-based regulatory systems provide access to large dynamic ranges that allow gene expression to be titrated to a level appropriate for the specific application. The dynamic ranges achievable in translational and posttranslational control systems are modest and higher basal expression levels are observed in the repressed state. Synthetic regulators that have both high dynamic range and low basal activity are still a challenge in the synthetic biology, especially in mammalian cells [161].

The advantages of transcriptional regulation are that there is remarkable flexibility in the design space for promoter sequences and DNA binding proteins and that known modules can be composed into larger systems. The response of a promoter to different inputs can be regulated by changing its repressor binding sites, introducing point mutations within its sequence or screening randomized libraries. Additionally, transcription can be controlled with synthetic TAL effector and zinc finger proteins as well as with the CRISPR Cas9 system or small transcription activating RNAs [162, 163, 164, 165, 166]. However, a transcriptional regulation is limited due to the loading effects, limited number of orthogonal regulators, and the time that it takes to process each step of gene expression. Moreover, some of the used transcriptional repressors often modified the chromatin structure surrounding the synthetic construct, and consequently epigenetic effects may lead to slow reactivation of many transcriptional regulators. While bacterial promoters are relatively easy to manipulate, their mammalian counterparts require more complex transcriptional machinery that varies among different cell types and states, which is limiting the generality and practical utility of synthetic promoters. Notwithstanding, synthetic promotors may also lead to metabolic burden and pleiotropic effects as a consequence of unnecessary overexpression [167, 168, 169 170].

Recently, there has been an increased effort to develop RNA based post-transcriptional tools to indirectly influence translation via mRNA stability and splicing

or to directly control translation by affecting ribosomal scanning. Most of post-transcriptional events such as alternative splicing or other RNA-based mechanisms like miRNAs, are often found in the lower dynamic range [171]. The low regulation from 2- to 4-fold also was observed in the RNA-based systems that interfere with the splicing mechanisms, although it was sufficient to trigger the cell death [61, 74]. Interestingly, the developed systems in these studies for splicing and translation control with the TetR aptamer shows a higher dynamic range up to 11-fold and low basal activity, what it represent a considerable improvement of the existing RNA regulators on the post-transcriptional level in human cells.

The translational regulation become an attractive approach used in mammalian gene regulation with various advantages over the transcriptional control. It allows for independent regulation from those sequence elements that are responsible for regulation at the DNA level. Once mRNA is transcribed, the genomic context surrounding the synthetic construct becomes irrelevant. Next, the dynamics of translational regulation are faster due to the skipping of the transcription step in gene expression and may be used for regulating processes on the medium timescale, such as feedback regulation to balance protein levels in therapeutic or metabolic circuits. Moreover, translational regulation can be directly interfaced with various cytoplasmic components of the cell. RNA aptamers may directly be utilized and incorporated into proteins or small molecule responsive translational regulators [162, 169].

A decrease of the metabolic burden can be achieved if RNA regulators are used instead of protein regulators of either transcription or translation. Production of RNA is cheaper in terms of nutrient requirements and does not divert ribosomal resources away from the translation of other cellular proteins. However, the RNA binding proteins-based devices contribute to the synthetic design by enhancing the regulatory effects compare to the small molecule responsive RNA switches [72].

As synthetic systems become more sophisticated, they will adapt a combination of transcriptional and various post-transcriptional regulators. There are an increasing number of new tools for translational regulation with a potency for future applications in mammalian synthetic biology [162, 172].

6 Material and Methods

Plasmid construction

Plasmids were constructed by standard cloning techniques using overlap extension or fusion PCR with Q5 Polymerase (NEB) and restriction and ligation reactions with HF restriction enzymes (NEB) and T4-DNA ligase (NEB), respectively. Custom oligonucleotides were synthesised by Sigma Aldrich. First, the *gfp* gene was cloned into pcDNA5/FRT vector, under CMV promotor, using the unique restriction sites for BamHI and NotI, resulting in EGFP* construct. Next, first the exon from CI4T5 construct, the complete intron with TetR aptamer and six nucleotides from the second exon were cloned into EGFP* by unique restriction sites for SpeI and MluI, resulting in the construct SP. All modifications of SP were done by overlap extension PCR and amplicon insertion by unique restriction sites for SpeI and MluI. Further, the mCherry or *MAX* genes were cloned into SP_cr by unique restriction sites for BamHI and NotI, resulting in the TSD_mCherry and TSD_MAX-GFP. Next, the MINX intron with flanking exon sequences [127] was placed in front of *egfp* reporter gene, generating MINX-EGFP. This construct was used for insertion the TetR aptamer resulting in A3SS-T1 and A3SS-T2. All modifications of A3SS-T2 were done by overlap extension PCR and amplicon insertion by unique restriction sites for SpeI, MluI, AgeI or KpnI. Furthermore, EGFP* construct was used to insert the TetR within 5' UTR resulting in T1-T3 constructs and then all modifications of T2 were done by overlap extension PCR and amplicon insertion by unique restriction sites for BamHI and NotI.

TetR protein was express under the CMV promoter and was modified at the N-terminus with a nuclear localization signal from c-MYC. The construct expressing TetR-mCherry was cloned by insertion of the mCherry gene into TetR expressing plasmid with unique restriction sites for KpnI and AgeI.

Table 6.1 Used the nuclear (NLS) and mitochondrial (MLS) localisation signals and 'self-cleavage' sequence (P2A).

Sequence name	Sequence (5' - 3')
NLS: c-MYC	CCGGCCGCGAAACGCGTGAAACTGGAT
MLS: COX8A	ATGTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCTCGGCCCGG CGGCTCCCAGTGCCGCGCGCCAA GATCCATTCGTTG
MLS: GLS	ATGATGCGGCTGCGAGGCTCGGGGATGCTGCGGGACCTGCTCCTGCGG
P2A	GGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAG GAGAACCCTGGACCT

Table 6.2 PCR mixture

Reagent	volume [µL]
Q5® reaction buffer (x5)	20
primer F (10 µM)	3
primer R (10 µM)	3
dNTPs (10 mM)	2
template (100 ng/µL)	1
Q5® DNA Polymerase (2 U/µL)	1
H_20	ad 100

Table 6.3 Overlap extension PCR program

Step	Temperature [°C]	Time	
Initial denaturation	98	30 s	
Denaturation	98	15 s	x35
Annealing	56-62	20 s	
Extension	72	20 s/kb	
Final extension	72	5 min	

Table 6.4 Fusion PCR program

Step	Temperature [°C]	Time
Denaturation	98	2 min
Annealing	58	2 min
Extension	72	30 min

Table 6.5 PCR colony mixture

Reagent	volume [µL]
Taq DNA reaction buffer (x10)	2.50
primer F (10 µM)	0.75
primer R (10 µM)	0.75
dNTPs (10 mM)	0.75
bacterial cell	1.00
Taq DNA Polymerase (5 U/µL)	0.25
H_20	ad 25.00

Table 6.6 PCR colony program

Step	Temperature [°C]	Time	
Initial denaturation	94	2 min	
Denaturation	94	30 s	x30
Annealing	54	30 s	
Extension	72	45 s	
Final extension	72	5 min	

Table 6.7 Digestion mixture

Reagent	volume [µL]
Cut Smart buffer (x10)	5
restriction enzyme I (0.4-0.8 U/µl)	1
restriction enzyme II (0.4-0.8 U/µl)	1
DNA (2-5 µg)	10
H_20	ad 50
	Incubation: 37°C; 1-2 h

Table 6.8 Ligation mixture

Reagent	volume [µL]
T4 DNA ligase buffer (x10)	2
backbone (25 ng)	1
insert (x ng)	1
ATP (10 mM)	2
T4 DNA ligase (400 U/µl)	1
H_20	ad 20
	Incubation: RT; 1-2 h

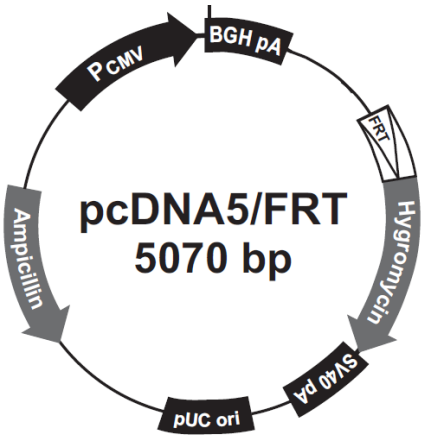


Figure 6.1 Plasmid map of pcDNA5/FRT. All constructs were cloned into pcDNA5/FRT under CMV-promoter (Invitrogen).

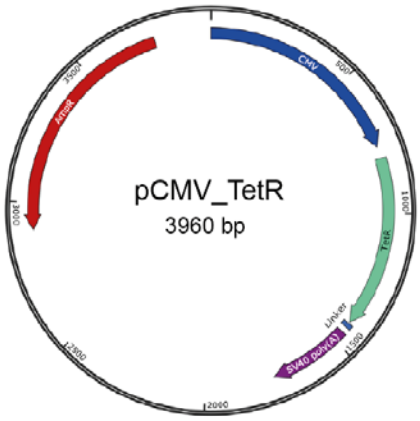


Figure 6.2 Plasmid map of CMV_TetR. Plasmid expressing under the CMV promoter the TetR protein. Based on this plasmid was cloned the TetR tagged with mCherry and additional plasmid expressing TetR-mCherry with NLS.

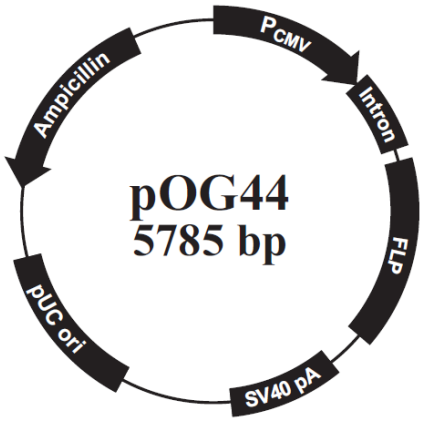


Figure 6.3 Plasmid map of pOG44. Plasmid expressing the Flp-recombinase and was used for stable integration (Invitrogen).

Table 6.9 Primers used for cloning.

Primer name	Sequence (5' - 3')
EGFP*_F	CACAGGATCCACTAGTATGACAAGTTCAACGCGTGTGAGCAAGGGCGAGGAGCTGTTCACCGGG
EGFP*_R	CTCGAGCGGCCGCTTACTTGTACAGCTCGTCCATGCCGAGAG
SP_F	GAAGATGCCAAAAACATTAAGAAGGG
SP_R	TCACACGCGTCGCTGGCTGTGGAGAGAAAGGC
SP_cr_F	ATCCACTAGTATGGAAGATGCCAAAAACATTAAGAAGGGCGTAAGTCTATCAAGCAGCATGTTATGG GTCATCACAGACCAGAGAAAAGCTTG
SP_cr_1_F	ATCCACTAGTATGGAAGATGGCAAAAACATTAAGAAGGCCGTAAGTCTATCGCAGCATGTTATGGGT CATCACAGACCAGAGAAAAGCGATAGATCAAGGTTACAAG
SP_cr_2_F	ATCCACTAGTATGGAAGATGGCAAAAACATTAAGAAGGCCGTAAGTCTAGCAGCATGTTATGGGTCA TCACAGACCAGAGAAAAGCTAGATCAAGGTTACAAG
SP_cr_3_F	ATCCACTAGTATGGAAGATGGCAAAAACATTAAGAAGGCCGTAAGTCGATCCCGCAGCATGTTATG GGTCATCACAGACCAGAGAAAAGCGGGATCGATCAAGGTTACAAG
SP_cr_4_F	ATCCACTAGTATGGAAGATGCCAAAAACATTAACAAGGGCGTAAGTG
SP_cr_5_F	ATCCACTAGTATGGAAGATGGCAAAAACATTAAGAAGGCCGTAAGTG
SP_cr_6_F	ATCCACTAGTATGGAAGATGGCAAAAACATTAACAAGGCCGTAAGTG
SP_cr_7_F	ATCCACTAGTATGGAAGATGCCAAAAACAGGAAGAACGGCGTAAGTCTATCAAGCAGCATG
SP_cr_8_F	ATCCACTAGTGAAGATCATCAAAAACATTAAGAAGGATGGTAAGTGTATCAAG
SP_cr_9_F	ATCCACTAGTGAAGATGGCAAAAACATTAACCATGGCCGTAAGTC
SP_cr_10_F	ATCCACTAGTGAAGATGCCAAAAACGCCACCATGGGCGTAAGTC
SP_cr_11_F	ATCCACTAGTATGATTAAGAAGGGCGTAAGTCTATCAAGCAG
SP_cr_12_F	ATCCACTAGTATGGAAGACGCCAAAAACATTAAGAAG
SP*_F	CAGCGACGCGTGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAG AACCCTGGACCTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGC
TSD_mCherry_F	CCAGCGACGCGTGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGG AGAACCCTGGACCTGTGAGCAAGGGCGAGGAGGATAACATG
TSD_mCherry_R	TGTGGCGGCCGCCTACTTGTACAGCTCGTCCATGCCGCCGGTGGAGTG
TSD_mCherry-1_F	CCAGCGACGCGTGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGG AGAACCCTGGACCTGTGAGCAAGGGCGAGGAGGATAACATAGCCATCATCAAGGAG
MAX-GFP_F	CCAGCGACGCGTGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGG AGAACCCTGGACCTAGCGATAACGATGACATCGAGGTGG
MAX-GFP_R	GCCCCAAAGCAGGAAGAAGCTCCGGATGGAGGCCAGCTCAGGAGGCGGTGGATCTGTGAGCAAG GGCGAGGAGCTGTTCACCGG
TetR-mCherry_F	GCGGCGACTCAGGAGGCGGTGGATCTGGTACCGTGAGCAAGGGCGAGGAGGATAAC
TetR-mCherry_R	CATCACCGGTTTACTTGTACAGCTCGTCCATGCCG
TetR-KpnI_F	CCCCGAATTCACCATGCCGG
TetR-KpnI_R	CATGTTATCCTCCTCGCCCTTGCTCACGGTACCAGATCCACCGCCTCCTGAGTCG
EGFP_NotI_R	GTGTGGCGGCCGCTTACTTGTACAGCTCGTCCATGCCGAGAGTGATCC
T1_F	CACACAGGATCCGCGTGTTACAGCATGTTACGGGTCATCACAGACCAGAGAAAAGTAACACGCGG AAGCGGAGCTACTAACTTCAGCCTGCTG
T2_F	CACACAGGATCCGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGG CGCGCGGAAGCGGAGCTACTAACTTCAGCCTGCTG
T3_F	CACACAGGATCCGCGTGTTACAGCATGTTATGGGTCATCACAGACCAGAGAAAAGTAACA CGCGGAAGCGGAGCTACTAACTTCAGCCTGCTG
T4_F	CACACAGGATCCCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCG CGGGGAAGCGGAGCTACTAACTTCAGCCTG
T5_F	CACACAGGATCCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGC GGGGAAGCGGAGCTACTAACTTCAG

T6_F	CACACAGGATCCCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGGG AAGCGGAGCTACTAACTTCAGCCTG
T7_F	CACACAGGATCCGCCGCAGCATGTTACGGGTCATC
T8_F	CACACAGGATCCCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGGGGGAA GCGGAGCTACTAAC
T9_F	CACACAGGATCCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGGAAGCGG AGCTACTAACTTCAG
T10_F	CACACAGGATCCGCGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAG CGGCGCGCGCGGGGAAGCGGAGCTACTAACTTCAGC
T11_F	CACACAGGATCCGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGG CGCGCGCGGGAAGCGGAGCTACTAACTTCAGC
T12_F	CACACAGGATCCGCGCGAAGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCTTCGCGCG GAAGCGGAGCTACTAACTTC
T13_F	CACACAGGATCCTCCCAGATCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGATCTGGGAGG AAGCGGAGCTACTAACTTCAGCCTGCTG
T14_F	CACACAGGATCCTCCCGGATCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGATCCGGGAGG AAGCGGAGCTACTAACTTCAGCCTGCTG
T15_F	CACACAGGATCCAGGGGATCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGATCCCCTGGGA AGCGGAGCTACTAACTTCAGCCTGCTG
T16_F	CACACAGGATCCTCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGCGAG GAAGCGGAGCTACTAACTTCAGCCTGCTG
T17_F	CACACAGGATCCTCGCGCTGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCAGCGCGAGG AAGCGGAGCTACTAACTTCAGCCTGCTG
T18_F	CACACAGGATCCTCGTGCTGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCAGCACGAGG AAGCGGAGCTACTAACTTCAGCCTGCTG
T19_F	CACACAGGATCCAGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCTGGAAGCG GAGCTACTAAC
T20_F	CACACAGGATCCACGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGTGGGGAAGCGGA GCTACTAAC
T21_F	CACACAGGATCCTTTGCGTGTTACAGCATGTTACGGGTCATCACAGACCAGAGAAAAGTAACACG CAAAGGAAGCGGAGCTACTAACTTCAGCCTG
T22_F	CACACAGGATCCGCGCGCTGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGC
T23_F	CACACAGGATCCGCGCGCTGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGTGCG GAAGCGGAGCTACTAAC
T24_F	CACACAGGATCCTGCCGCAGCATGTTACGGGTCATC
T25_F	CACACAGGATCCGCGCGACGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCG
T26_F	CACACAGGATCCGCGCGCCAGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGC
T27_F	CACACAGGATCCGCGCGCCAGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGACG CGGGGAAGCGGAGCTACTAACTTC
T28_F	CACACAGGATCCGCGCGCCGCAGCATGTTACGGGCCGAAAGGCCAGAGAAAAGCGGCGCGCGG GGAAGCGGAGCTACTAACTTC
T29_F	CACACAGGATCCGCGCCGCAGCATGTTACGGGCCGAAAGGCCAGAGAAAAGCGGCGCGGGAAG CGGAGCTACTAACTTCAGCCTG
T30_F	CACACAGGATCCGCGCGCCGCAGCATGTTACGGGCCTTCGGGCCAGAGAAAAGCGGCGCGCGG GGAAGCGGAGCTACTAACTTC
T31_F	CACACAGGATCCGCGCCGCAGCATGTTACGGGCCTTCGGGCCAGAGAAAAGCGGCGCGGGAAGC GGAGCTACTAACTTC
T32_F	CTAGAGTACTCGTACGCACACAGGATCCGGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGG CTGGAGACG
T32_R	CGTCTCCAGCCTGCTTCAGCAGGCTGAAGTTAGTAGCTCCGCTTCCGGATCCTGTGTGCGTACGAG TACTCTAG
T33_F	CACACAAAGCTTGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGCGCG ATTCTTCTGACACAACAGTCTCGAACTTAAGC
T33_R	GTGAACAGCTCCTCGCCCTTGCTCACCATGGATCCTGTGTGCGTACGAGTACTCTAGCCTTAAG
T33_2F	CTTAAGGCTAGAGTACTCGTACGCACACAGGATCCATGGTGAGCAAGGGCGAGGAGCTGTTCAC
T2*_F	CACACAGGATCCGCGCGCCGCAGTATGTTACAGGTCATCACAGACCAGAGAAAAGC
T6*_F	CACACAGGATCCCGCCGCAGTATGTTACAGGTCATCACAGACCAG
TT1_F	AGTACTCGTACGGACACAGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGC GCGCGGAGAGGCG
TT1_R	GCGCGCGGATCCGCCTCCGCGCTTCTTAATGTTTTTCGCCTCTCCGCGCGCCGCTTTTCTCTGGTC
TT2_F	GGAGGCGGATCCGCGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCGCG CGGGAAGCGGAGCTACTAACTTC
TT3_R	GTGTGCGTACGAGTACTCTCGCCTTCAGAGCTGTAATTGAACTGGGAGTGGACACCTGCCCAGTGC CTCGACCAACTTCTG
TT4_F	TGTGTGCGTACGAGTACTCTCGCCTTCAGAGCTGTAATTAAGCGCGCCGCTTTTCTCTGGTCTGTGA TGACCCGTAACATGCTGCGGCGCGCAACTTCTGCAGCTTAAGTTCGAGAC
del_NLS_TetR_F	GGCCCCGAATTCACCATGTCTAGACTGGACAAGAGCAAAGTCATAAAC

T2-SP*_F	CACACAGGATCCGCGCGCCGCAGCATGTTACGGGTCATCACAGACCAGAGAAAAGCGGCG CGCGGAAGCGGAGCTACTAACTTCAGCCTGCTG
MINX_F	ATCCGGAAACCCGTCGGCCTCCGAACGGTAAGAGCCTAGCATGTAGAACTGGTTACCTGCAGCCCAA GCTTGCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCCTTG
MINX_R	ATCACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAAAGGGACAGGATAAGTATGTCGG TACCTCATCAAGGAAACCCTGGACTACTGCGCCCTAGAC
MINX-EGFP_F	ATCCACTAGTATGCGGAAACCCGTCGGCCTCCGAAC
MINX-EGFP_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACC
A3SS-T1_F	GATGAGGTACCGACATACTTATCCTGTCCCTTTTGCATGTTACAGCATGTTATGGGTCATCACAGACCAG AGAAAAGTAACATGC
A3SS-T1_R	ACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATGTTACTTTTCTCTGGTCTGT GATGACCC
A3SS-T2_F	GATGAGGTACCGACATACTTATCCTGTCCCTTTTGCATGTTACAGCATGTTATGCCTCATCACCGAGGAG AGAAAAGTAACATGC
A3SS-T2_R	ACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATGTTACTTTTCTCTCCTCGGTGATGA GGC
A3SS-T3_F	GATGAGGTACCGACATACTTATCCTGTCCCTTTTGCGCCAGCATGTTATGGGTCATCACAGACCAGAGA AAAGGCGC
A3SS-T3_R	ACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCGCCTTTTCTCTGGTCTGTGAT GACCC
A3SS-T4_F	ATCCACTAGTCGGAAACCCGTCGGCCTCCGAACGATGGTAAGAGCCTAG
A3SS-T4_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATGCCACTTTTCTCTGG TCTGTGATGACCCATAACATGCTGTGGCATGCAAAAGGGACAGG
A3SS-T5_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATGTTCCTTTTCTCTGG TCTGTGATGACCCATAACATGCTGGAACATGCAAAAGGGACAG
A3SS-T6_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGGCATGTTACTTTTCTCTG GTCTGTGATGACCCATAACATGCTGTAACATGCCAAAAGGGACAGGAT
A3SS-T7_F	GATGAGGTACCGACATACTTATCCTGTCCCTTTTGCATTGTGTTACAGCATGTTATGGGTCATCACAGAC CAGAGAAAAGTAACACGATGC
A3SS-T7_R	GCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATCGTGTTACTTTTCTCTGGTCT GTGATGACCC
A3SS-T8_F	GATGAGGTACCGACATACTTATCCTGTCCCTTTTGCATGTTCCAGCATGTTATGGGTCATCACAGACCAG AGAAAAGGAACATGC
A3SS-T8_R	CGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAAAGCATGTTCCTTTTCTCTGGT CTGTGATGACCC
A3SS-T9_F	GATGAGGTACCGACATACTTATCCTGTCACGATCGCATGTTCCAGCATGTTATGGGTCATCACAGACCA GAGAAAAGGAACATGC
A3SS-T9_R	ACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAGCATGTTCCTTTTCTCTGGTCTGTGATGAC CC
A3SS-T10_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTTTTCCGGAAAAGGCATGTTACTTTT CTCTGGTCTGTGATGACC
A3SS-T11_R	CTCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAGGCATGTTACTTTTCT CTGGTCTGTGATGACCCATAACATGCTGTAACATGCCAAAGGACAGGATAA
A3SS-T11*_R	CTCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAGGCATGTTACTTTTCT CTGGTCTGTGATGACCTTTAACATACTGTAACATGCCAAAGGACAGG
A3SS-T12_R	CTCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAAAGGCATGTTACTTTT CTCTGGTCTGTGATGACCCATAACATGCTGTAACATGCCAAAAGGGACAGG
A3SS-T13_F	GATGAGGTACCGACATACTTATCCTGTCTTTTGCACGTTACAGCATGTTATGGGTCATCACAGACCAGAG AAAAGTAACGTGC
A3SS-T13_R	GCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAAAAAAAGCACGTTACTTTTCTCT GGTCTGTGATGACCC
A3SS-T14_R	ACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAGGCACGCCACTTTTCTCTGG TCTGTGATGACCCATAACATGCTGTGGCGTGCCAAAGGACAGGATAAG
A3SS-TN1_F	CTTGCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGAGGTACCGACATACTTATCCTGTCC TTTGGCATGTTACAGCATGTTATGGATCCGGCCGCGAAACG
A3SS-TN1_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGGGAAAAAAGGCATGTTACTTTTCTCT GATCCAGTTTCACGCGTTTCGCGGCCGGATCCATAACATGCTGTAACA

A3SS-TN2_F	GATGAGGTACCGACATACTTATCCTGTCCTTTGGCGGTTACAGCATGTTATGGATCCGGCCGCGAAACG CGTGAAACTGGATC
A3SS-TN2_R	CGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGTGGAAAAAAGGCAGTTACTTTTCTCTGATCC AGTTTCACGCGTTTCGCGGCCGGATCCAT
A3SS-TN3_F	CTTGCTGCACGTCTAGGGCGCAGTAGTCCAGGGTTTCCTTGATGAGGTACCGACATACTTATCCTGTCC TTTGGCATGTTACAGCATGTTATGGATCCGGCCGCGAAACG
A3SS-TN3_R	TCACACGCGTACCGGTGAAGAGTTTGTCCTCAACCGCGAGCTGGGAAAAAAGGCATGTTACTTTTCTCT GATCCAGTTTCACGCGTTTCGCGGCCGGATCCATAACATGCTGTAACA
A3SS-TN2-CP_F	GATCTGGTACCCCTGGGCACTTACAGGAAGGCTTCGGCTGCGTGGTCACC
A3SS-TN2-CP_R	GTGTGGCGGCCGCTTAAGCCAGAGCTGGGAATGCCTCTGGGTCATCCACATC
A3SS-TN2-CP_2F	GATGTGGATGACCCAGAGGCATTCCCAGCTCTGGCTTCAGGAGGCGGTGGATCTGTGAGCAAGGGCGA GGAGCTGTTCACCGG
A3SS-TN2-CP_2R	CCGGTGAACAGCTCCTCGCCCTTGCTCACAGATCCACCGCCTCCTGAAGCCAGAGCTGGGAATGCCT CTGGGTCATCCACATC
MLS-EGFP_F	GGATCCACTAGTATGTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCTCGGCCCGGCGGC TCCCAGTGCCGCGCGCCAAGATCCATTCGTTGACGCGTGTGAG
NLS-MLS-EGFP_F	CACACAGGATCCACTAGTATGCCGGCCGCGAAACGCGTGAAACTGGATCTCGCGGTTGAGGACAAAC TCTTCTCCGTCCTGACGCCGCTGCTGCTGCGGGGCTTGACAGGCTC
A3SS-TN2-MLS_F	GAGAAAAGTAACTGCCTTTTTTCCACAGCTCGCGGTTGAGGACAAACTCTTCACCGGTTCCGTCCTGAC GCCGCTGCTGCTGCGGGGCTTGACAGGCTCGGC
A3SS-TN2-MLS_R	GCCGAGCCTGTCAAGCCCCGCAGCAGCAGCGGCGTCAGGACGGAACCGGTGAAGAGTTTGTCCTCAA CCGCGAGCTGTGGAAAAAAGGCAGTTACTTTTCTC
MLS-GLS-EGFP_F	CTCTACAACTGTAGTATATAGAATGGAGAGTCTGGGGGAGAGGAGCGGCGACTCAGGAGGCGGTGGAT CTACCGGTGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG

RT-PCR and lariat PCR analysis

A total of 120,000 HeLa cells per well were seeded into a 12-well plate. The cells were transfected with 100 ng reporter DNA and 300 ng TetR plasmid using 2 µl of Lipofectamine® 2000 (Life Technologies) per well according to manufacturer instructions. Then, 2 h after transfection, the medium was changed to DMEM with or without 50 µM dox, and the cells were incubated at 37°C and 5% CO_2 for 24 h. RNA was isolated using TRIzol® (Life Technologies) according to the manufacturer's protocol. Contaminating DNA was removed with the TURBO DNA-free kit (Life Technologies), and the RNA was quality checked on a 1% (w/v) agarose gel. Next, 1 µg RNA was reverse-transcribed by MuLV (Applied Biosystems) using random hexamers (Fermentas) with the supplied buffers (10 min at 20°C, 15 min at 42°C, 5 min at 99°C). Then, 50 ng cDNA was PCR amplified using Taq polymerase (New England Biolabs, initial denaturation 2 min at 96°C, 30 sec at 96°C, 30 sec at 54°C, 30 sec at 72°C, 35 cycles) and analysed on a 3% (w/v) agarose gel. The amplified products were cloned (CloneJET PCR *Cloning Kit,* Thermo Scientific) and sequenced for verification. Each RT-PCR was repeated in three independent experiments. The same way cDNA used for lariat PCR was prepared but with the 20 cycles. Next products were diluted 1:20, re-amplified with nested primers, with 30 cycles and

analyzed on a 3% (w/v) agarose gel. Bands of interest using CloneJET were cloned and sequenced for verification. Primers are listed below.

Table 6.10 Used primers for RT-PCR and lariat PCR analysis.

Name	Sequence (5' - 3')
5SS_F	GTACGCACACAGGATCCACTAGTATG
5SS_R	GTGAACAGCTCCTCGCCCTTG
A3SS_F	AACCCGTCGGCCTCCGAAC
A3SS_R	CGCCGGACACGCTGAACTTG
lariat_1F	GTCTAGGGCGCAGTAGTCCAG
lariat_1R	GCAGCAAGCTTGGGCTGCAG
lariat_2F	CTAGGGCGCAGTAGTCCAGG
lariat_2F	GCTGCAGGTAACCAGTTCTAC

qPCR analysis

For qPCR analysis the Fast SYBR Green Master Mix (Applied Biosystems) was used and the samples were analysed on a StepOnePlus Real-Time PCR machine (Applied Biosystems) according to the supplier's protocol. Analysis was performed with samples from three independent experiments in technical replicates. Results according to the ΔΔCt method were calculated [173]. Primers are listed below.

Table 6.11 Used primers for qPCR analysis.

Name	Sequence (5' - 3')
spliced_1F	CATTAAGAAGGGCCCAGCGAC
unspliced_1F	CTTGCGTTTCTGATAGGCACC
GFP_1R	AGGGTTCTCCTCCACGTCTC
actin_F	CGGGACCTGACTGACTACCTC
actin_R	CTTCTCCTTAATGTCACGCACG
TetR_F	CGCTCAAAAGCTGGGAGTTG
TetR_R	GCCTGTCCAGCATCTCGATT

Table 6.12 qPCR mixture

Reagent	volume [µL]
SYBR Green Master Mix (x2)	10
Primer Mix (each 10 µM)	1
cDNA	3
H_20	ad 25

Table 6.13 qPCR program

Temperature [°C]	Time	
95	20 s	x40
95	3 s	
60	30 s	
95	15 s	melting curve program
60	60 s	
95	15 s	

Agarose gel electrophoresis

The analysis of size and quality of DNA and RNA as well as the DNA fragment purification was performed through applying agarose gel electrophoresis. Dependent on fragment size, 1% - 3% (w/v) agarose gels were used. For their fabrication, the respective amount of agarose was dissolved in 1x TAE and dissolved by heating in the microwave. The samples were mixed with 6x DNA or 2x RNA loading dye, the gel was run in 1x TAE and an electric field strength of 6 V/cm. Gels were stained in 0.5 µg/ml ethidium bromide for 15 min and DNA or RNA was visualized under UV light.

Table 6.14 Composition of 50xTAE and 6x DNA loading dye.

Buffer/ solution	Ingredients	Concentration
50x TAE	Tris	2 M
	Acetic acid	1 M
	EDTA	50 mM
	pH 8.3	
6x DNA loading dye	Tris-HCl pH 7.6	40 mM
	EDTA	1 mM
	Acetic acid	20 mM
	Glycerol	50% (v/v)
	Bromophenol blue	spatula point
	Xylene cyanole	spatula point

DNA precipitation with butanol

To concentrate DNA, 10 volumes of 1-butanol were added to the DNA solution and the mix was vortexed for 20 s. Then, it was centrifuged for 15 min at 13.000 rpm, washed with 75% (v/v) ethanol and centrifuged for 10 min at 13.000 rpm, again. After removing the supernatant, the pellet was dried for 5 min at RT and dissolved in 8 µl water.

DNA purification

DNA purification was conducted by using column purification or gel extraction kits. To purify DNA from proteins, salt, RNA and small DNA fragments, the QIAquick PCR Purification kit (QIAGEN) was used. To isolate DNA fragments of defined size, the DNA sample was run on a gel and the band of interest was excised from the gel. Gel extraction was performed employing the QIAquick Gel Extraction kit (QIAGEN) according to the manufacturer`s protocol. The pellet was dissolved in water.

Plasmid preparation

For the propagation of a plasmid in bacteria, the cells containing the desired plasmid were inoculated in 4 ml of LB-medium supplemented with the respective antibiotic and grown at 37°C shaking at 150 rpm overnight. For plasmid isolation the QIAprep®Spin Miniprep kit (QIAGEN) was used according to the manufacturer`s protocol. The pellet was dissolved in water.

Table 6.15 Composition of ampicillin stock and LB and SOC medium.

Buffer/ solution	Ingredients	Concentration
Ampicillin (stock solution)	Ampicillin in 70% (v/v) EtOH	100 mg/ml
LB medium	Tryptone	1% (w/v)
	Yeast extract	0.5% (w/v)
	NaCl	1% (w/v)
	Ampicillin	100 µg/ml
LB-amp plates	Agar	2% (w/v)
SOC medium	Yeast Extract	0.5%
	Tryptone	0.2%
	NaCl	10 mM
	KCl	2.5 mM
	$MgCl_2$-7·H_20	10 mM
	$MgSO_4$-6·H_20	10 mM

Determination of the concentration of nucleic acids

The concentration of DNA and RNA sample was determined at 260 nm wave length using the NanoDrop® ND-1000 Spectrophotometer. As reference the respective solvent was employed.

Preparation of electrocompetent cells

One day before starting the preparation, two overnight cultures of 4 ml LB-medium containing tubes were inoculated with *E.coli* Top10 cells (Invitrogen) and grown at 37°C at 150 rpm overnight. Next day, the cultures were transferred to 600 ml of LB-medium in a 2 L baffle flask to yield a start OD600 of 0.1. The cell suspension was incubated at 37°C at 200 rpm. After reaching an OD600 between 0.6 - 0.8 cells were chilled on ice for 0.5 h, harvested by centrifugation for 20 min at 7.000 rpm and washed three times with 10% (v/v) of glycerol. Cells were resuspended in 4 ml of 10% (v/v) of glycerol and split into aliquots of 100 µl in 1 ml reaction tubes.

Electroporation of *E.coli*

Electroporation was used to transform *E.coli* cells. 0.5 µl of DNA were gently mixed with the bacteria and the mixture transferred into the slit of the micro pulser electroporation cuvette, 0.1 cm gap (Bio-Rad). The bacteria were electro-shocked with 1.800 V for 5 ms. Immediately after applying the electric field to the cells, cells were extracted from the cuvette and mixed with 450 µl SOC-medium. After a 45 min recovery phase at 37°C and 1.000 rpm in the shaker, 200 µl of cells were plated onto agar plates containing ampicillin. Plates were incubated at 37°C overnight.

Cell culture

HEK293 and HeLa cells (DSMZ, No. ACC-305 and ACC-57, respectively) and HF1-3 cells “Flp-In Host Cell Line” [174] were maintained at 37°C in a 5% CO_2 humidified incubator and cultured in Dulbecco’s Modified Eagle Medium (DMEM, Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS Superior, Biochrom), 100 U/ml penicillin (PAA, the Cell Culture Company), 100 µg/ml streptomycin (PAA, the Cell Culture Company) and 1 mM sodium pyruvate (PAA The Cell Culture Company). For HF1-3 cells, 150 µg/mL zeocin (Invitrogen) was additionally supplemented to the medium, whereas the medium of the HF1-3 cells harbouring the integrated constructs P2A_GFP, TSD_GFP, A3SS-TN2-CP and A3SS-TN2-MLS contained 200 µg/mL hygromycin (Invitrogen).

Genomic Integration

HF1-3 cells were transfected with the plasmids pcDNA5/FRT_GFP and pOG44 (recombinase expression plasmid, Invitrogen) at a molar ratio of 1:9 using Lipofectamine® 2000 (Life Technologies) according to manufacturer instructions. The medium was changed 24 h after transfection to DMEM. The cells were selected for stable integration by adding 200 µg/ml hygromycin. After two weeks of selection with hygromycin, cells were sorted by the S3e™ Cell Sorter (Bio-Rad) for EGFP positive cells and analysed for stable integration by genomic PCR and sequencing.

Western blotting

For western blot analyses, cells were transfected with the respective plasmids and grown over night with or without dox. Cells were lysed in T-PER™ tissue protein extraction reagent (Life Technology) with EDTA-free protease inhibitor (Roche) for 30 min at 4°C. Protein concentration was measured by the Bradford method (Bio-Rad) according to the instructions provided by the manufacturer. 10 µg protein lysate was separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis 12 %, then transferred to a HyBond ECL nitrocellulose membrane (Amersham) and blocked with Odyssey blocking buffer (LI-COR® Bioscience) for 1h at RT. The membranes were next incubated 1.5 h with primary antibodies that recognize GFP (Sigma-Aldrich) and actin (Santa Cruz Biotechnology) in 1:1000 and 1:7000 dilution, respectively. Membranes were washed with PBS (Invitrogen) pH 7.4/Tween20 0.1% (v/v) (Carl Roth) and incubated for 45 min at RT with infrared dye conjugated secondary antibodies IRDye® 800CW Donkey anti-Mouse IgG (Santa Cruz Biotechnology) and IRDye® 800CW Donkey anti-Goat IgG (Santa Cruz Biotechnology) for GFP and actin, respectively. Then membranes were washed with PBS-Tween buffer. Visualisation was carried out with Odyssey Infrared Imaging System (Image Studio Lite; LI-COR® Bioscience. Three independent experiments were performed.

Flow Cytometry

A total of 120,000 HEK293 or 60,000 HeLa cells per well were seeded into a 24-well plate. The cells were transfected with 50 ng reporter DNA and 150 ng TetR plasmid using 1 µl Lipofectamine® 2000 (Life Technologies) per well according to manufacturer instructions. The medium was changed 2 h after transfection to DMEM (without phenol red) with or without 50 µM doxycycline (Sigma-Aldrich). The GFP and mCherry

expression was measured after 24 h using flow cytometry (CytoFlex S, Beckman Coulter). Mean values and standard deviations were calculated from triplicates and normalised to the TetR-mCherry. Each experiment was repeated three times. GFP and mCherry were excited using a 488 nm laser and a 550/30 filter and 561 nm laser and a 610/20 filter, respectively. Cells were analysed using FlowJo (TreeStar Inc., Ashland, OR), and populations were selected by gating out the GFP background signal of untransfected cells. In the case of stable integrated cells, a total of 60,000 HF1-3 cells expressing P2A_GFP control or TSD_GFP construct per well were seeded into a 12-well plate. The cells were transfected with 300 ng TetR plasmid using 2 µl Lipofectamine® 2000 (Life Technologies) per well according to manufacturer instructions. The medium was changed 4 h after transfection to DMEM (without phenol red) with or without 50 µM doxycycline. The GFP and mCherry expressions were measured after 72 h using flow cytometry. Populations were selected by gating out the GFP background signal of untransfected cells with TetR-mCherry.

Microscopic images

Hela or HF1-3 cells expressing EGFP or A3SS-T2-CP grown on glass coverslips for 24 h and next were transfected with 300 ng TetR-mCherry plasmid using 2 µl of Lipofectamine® 2000 (Life Technologies) according to the manufacturer's instructions. Then, 2 h after transfection, the medium was changed to DMEM with or without 50 µM dox, and the cells were incubated at 37°C and 5% CO2 for 24 h. Further, cells were fixed in 3,7% formaldehyde at room temperature (RT) for 10 minutes followed by permeabilization with 0.5% triton X-100 in 1xPBS for 10 minutes at RT. Cells were then washed three times with 1xPBS and stained with 4′,6-diamidino-2-phenylindole (DAPI) dye (1 µg/mL in water; Thermo Scientific) and washed three times with 1xPBS. Next, coverslips were mounted with mounting medium (Thermo Scientific). All experiments were carried out on a Zeiss Axiovert 200 M inverted microscope using a 20x/0.5 NA or 63x/1.4 NA oil objective lens (Zeiss). Excitation was done using the mercury arc lamp (Zeiss). The filters used were 350/50 (excitation) and 460/50 (emission) for DAPI, 482/18 (excitation) and 520/28 (emission) for EGFP and 565/30 (excitation) and 620/60 (emission) for mCherry and mito tracker (ab176831, abcam). Images were repeated in three independent experiments. The data processing were performed using ImageJ.

7 Appendix

Table 7.1 List of abbreviations.

Abbreviation	Meaning
% (w/v)	% (weight/ volume)
°C	degrees Celsius
µg	microgramm
µL	microliter
µM	micromolar
3SS	3' splice site
5SS	5' splice site
5’ UTR	5' untranslated region
A	adenine
AS	alternative splicing
amp	Ampicilin
BP	branch point sequence
C	cytosine
cDNA	complementary DNA
CI	chimeric intron
CP	cytosolic protein
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
dNTP	desoxynucleoside triphosphate
EGFP	enhanced green fluorescent protein
FCS	fetal calf serum
FRT	Flp-In recombination target
F	forward
G	guanine
GOI	gene of interest
kb	kilobases
min	minutes
mL	mililiter
mM	milimolar
mRNA	messenger RNA
nm	nanometer
nM	nanomolar
nt	nucleotide
ORF	open reading frame
P2A	2A ‘self-cleaving’ peptide from the porcine teschovirus-1
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Polypyrimidine	Py

R	reverse
RNA	ribonucleic acid
rpm	rounds per minute
RT	room temperature
s	seconds
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
snRNPs	small nuclear ribonucleoproteins
T	thymine
TetR	tetracycline repressor
TSD	TetR Splicing Device
U	uracil
uORF	upstream open reading frame
UV	ultra violet

8 References

1. Qian, Y., McBride, C. & Del Vecchio, D. Programming Cells to Work for Us. *Annu. Rev. Control. Robot. Auton. Syst.* 1, 4.1-4.30 (2018).
2. Slomovic, S., Pardee, K. & Collins, J. J. Synthetic biology devices for in vitro and in vivo diagnostics. *Proc. Natl. Acad. Sci.* 112, 14429–14435 (2015).
3. Vecchio, D. Del, Dy, A. J. & Qian, Y. Control Theory Meets Synthetic Biology. (2016). doi:http://dx.doi.org/10.1098/rsif.2016.0380
4. Ding, Y., Wu, F. & Tan, C. Synthetic Biology: A Bridge between Artificial and Natural Cells. *Life* 4, 1092–1116 (2014).
5. Elowitz, M. B. & Leibier, S. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338 (2000).
6. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* 403, 339–342 (2000).
7. Vazquez-Anderson, J. & Contreras, L. M. Regulatory RNAs: Charming gene management styles for synthetic biology applications. *RNA Biol.* 10, 1778–1797 (2013).
8. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422 (2009).
9. Lewis, D. D., Villarreal, F. D., Wu, F. & Tan, C. Synthetic Biology Outside the Cell: Linking Computational Tools to Cell-Free Systems. *Front. Bioeng. Biotechnol.* 2, (2014).
10. Mathur, M., Xiang, J. S. & Smolke, C. D. Mammalian synthetic biology for studying the cell. *J. Cell Biol.* 216, 73–82 (2017).
11. Bird, A. *et al.* Multi-Input RNAi-Based Logic Circuit. 1307–1312 (2011).
12. Din, M. O. *et al.* Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* 536, 81–85 (2016).
13. Wei, P. *et al.* Bacterial virulence proteins as tools to rewire kinase pathways in yeast and immune cells. *Nature* 488, 384–388 (2012).
14. Chakravarti, D. & Wong, W. W. Synthetic biology in cell-based cancer immunotherapy. *Trends Biotechnol.* 33, 449–461 (2015).
15. Saxena, P. *et al.* A programmable synthetic lineage-control network that differentiates human IPSCs into glucose-sensitive insulin-secreting beta-like cells. *Nat. Commun.* 7, 1–14 (2016).
16. Chappell, J., Watters, K. E., Takahashi, M. K. & Lucks, J. B. A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Curr. Opin. Chem. Biol.* 28, 47–56 (2015).
17. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* 12, 381–390 (2014).
18. Groher, F. & Suess, B. Synthetic riboswitches - A tool comes of age. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1839, 964–973 (2014).
19. Hallberg, Z. F., Su, Y., Kitto, R. Z. & Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem* 86, 515–39 (2017).
20. Chappell, J. *et al.* The centrality of RNA for engineering gene expression.

Biotechnol. J. 8, 1379–1395 (2013).
21. Chen, Y. Y. *et al.* Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome Biol.* 13, 240 (2012).
22. McKeague, M., Wong, R. S. & Smolke, C. D. Opportunities in the design and application of RNA for gene expression control. *Nucleic Acids Res.* 44, 2987–2999 (2016).
23. Engreitz, J. M., Ollikainen, N. & Guttman, M. Long non-coding RNAs: Spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 756–770 (2016).
24. Dykes, I. M. & Emanueli, C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics, Proteomics Bioinforma.* 15, 177–186 (2017).
25. Kaphingst, K. A., Persky, S. & Lachance, C. NIH Public Access. 14, 384–399 (2010).
26. Beisel, C. L. & Storz, G. Networks. *FEMS Microbiol. Rev.* 34, 866–882 (2011).
27. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol.* 5, 1998–2010 (2007).
28. Seto, A. G., Kingston, R. E. & Lau, N. C. The Coming of Age for Piwi Proteins. *Mol. Cell* 26, 603–609 (2007).
29. Jusiak, B., Cleto, S., Perez-Piñera, P. & Lu, T. K. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends Biotechnol.* 34, 535–547 (2016).
30. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* 152, 17–24 (2013).
31. Berens, C., Groher, F. & Suess, B. RNA aptamers as genetic control devices: The potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* 10, 246–257 (2015).
32. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by bindingS-adenosylmethionine. *Nat. Struct. Mol. Biol.* 10, 701–707 (2003).
33. Wittmann, A. & Suess, B. Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* 586, 2076–2083 (2012).
34. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281–286 (2004).
35. Kim, J. N. & Breaker, R. R. Purine sensing by riboswitches. *Biol. Cell* 100, 1–11 (2008).
36. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 8, (2007).
37. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: Structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* 3, 1–13 (2011).
38. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* 4, 1–15 (2012).
39. Wachter, A. *et al.* Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3' End Processing of mRNAs. *Plant Cell Online* 19, 3437–3450

(2007).
40. Zhuo, Z. *et al.* Recent advances in SELEX technology and aptamer applications in biomedicine. *Int. J. Mol. Sci.* 18, 1–19 (2017).
41. Suess, B. & Weigand, J. E. Nucleic Acid and Peptide Aptamers. 535, 201–208 (2009).
42. Wachsmuth, M. *et al.* Design criteria for synthetic riboswitches acting on transcription. *RNA Biol.* 12, 221–231 (2015).
43. Topp, S. & Gallivan, J. P. Guiding Bacteria with Small Molecules and RNA Guiding Bacteria with Small Molecules and RNA. 129, 6807–6811 (2007).
44. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* 41, 2541–2551 (2013).
45. Ceres, P., Garst, A. D., Marcano-Velázquez, J. G. & Batey, R. T. Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. *ACS Synth. Biol.* 2, 463–472 (2013).
46. Ceres, P., Trausch, J. J. & Batey, R. T. Engineering modular 'ON' RNA switches using biological components. *Nucleic Acids Res.* 41, 10449–10461 (2013).
47. Yen, L. *et al.* Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 431, 471–476 (2004).
48. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl. Acad. Sci.* 104, 14283–14288 (2007).
49. Wieland, M. & Hartig, J. S. Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew. Chemie - Int. Ed.* 47, 2604–2607 (2008).
50. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA intera ... *Library (Lond).* 282, 296–298 (1998).
51. Grate, D. & Wilson, C. Inducible regulation of the S. cerevisiae cell cycle mediated by an RNA aptamer-ligand complex. *Bioorganic Med. Chem.* 9, 2565–2570 (2001).
52. Weigand, J. E. *et al.* Screening for engineered neomycin riboswitches that control translation initiation. *RNA* 14, 89–97 (2008).
53. Harvey, I., Garneau, P. & Pelletier, J. Inhibition of translation by RNA-small molecule interactions. *Rna* 8, 452–463 (2002).
54. Ogawa, A. Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. *Rna* 17, 478–488 (2011).
55. Bocobza, S. *et al.* Riboswitch-dependent gene regulation and its evolution in the plant kingdom Riboswitch-dependent gene regulation and its evolution in the plant kingdom. 2874–2879 (2007). doi:10.1101/gad.443907
56. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447, 497–500 (2007).
57. Croft, M. T., Moulin, M., Webb, M. E. & Smith, A. G. Thiamine biosynthesis in

algae is regulated by riboswitches. *Proc. Natl. Acad. Sci.* 104, 20770–20775 (2007).
58. Kim, D.-S., Gusti, V., Pillai, S. G. & Gaur, R. K. An artificial riboswitch for controlling pre-mRNA splicing. *RNA* 11, 1667–1677 (2005).
59. Kim, D.-S., Gusti, V., Dery, K. J. & Gaur, R. K. Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC Mol. Biol.* 9, 23 (2008).
60. Weigand, J. E. & Suess, B. Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res.* 35, 4179–4185 (2007).
61. Vogel, M., Weigand, J. E., Kluge, B., Grez, M. & Suess, B. A small, portable RNA device for the control of exon skipping in mammalian cells. *Nucleic Acids Res.* 1–12 (2018). doi:10.1093/nar/gky062
62. Ausländer, S., Ketzer, P. & Hartig, J. S. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Mol. Biosyst.* 6, 807 (2010).
63. Beilstein, K., Wittmann, A., Grez, M. & Suess, B. Conditional Control of Mammalian Gene Expression by Tetracycline-Dependent Hammerhead Ribozymes. *ACS Synth. Biol.* 4, 526–534 (2015).
64. Klauser, B., Atanasov, J., Siewert, L. K. & Hartig, J. S. Ribozyme-Based Aminoglycoside Switches of Gene Expression Engineered by Genetic Selection in S. cerevisiae. *ACS Synth. Biol.* 4, 516–525 (2015).
65. Nomura, Y., Zhou, L., Miu, A. & Yokobayashi, Y. Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS Synth. Biol.* 2, 684–689 (2013).
66. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11, 597–610 (2010).
67. Kurreck, J. RNA interference: From basic research to therapeutic applications. *Angew. Chemie - Int. Ed.* 48, 1378–1398 (2009).
68. Atanasov, J., Groher, F., Weigand, J. E. & Suess, B. Design and implementation of a synthetic pre-miR switch for controlling miRNA biogenesis in mammals. *Nucleic Acids Res.* 45, e181 (2017).
69. Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol. Syst. Biol.* 4, (2008).
70. Tuleuova, N., An, C. Il, Ramanculov, E., Revzin, A. & Yokobayashi, Y. Modulating endogenous gene expression of mammalian cells via RNA-small molecule interaction. *Biochem. Biophys. Res. Commun.* 376, 169–173 (2008).
71. Kumar, D., An, C. Il & Yokobayashi, Y. Conditional RNA interference mediated by allosteric ribozyme. *J. Am. Chem. Soc.* 131, 13906–13907 (2009).
72. Aragiannis, P., Ujita, Y. & Aito, H. RNA-based gene circuits for cell regulation. *Proc. Jpn. Acad.* 92, 412–422 (2016).
73. Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat. Biotechnol.* 33, 839–841 (2015).
74. Culler, S. J., Hoff, K. G. & Smolke, C. D. Reprogramming cellular behavior with

RNA controllers responsive to endogenous proteins. *Science (80-.).* 330, 1251–1255 (2010).

75. Wang, Y., Cheong, C. G., Tanaka Hall, T. M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat. Methods* 6, 825–830 (2009).
76. Cooke, A., Prigge, A., Opperman, L. & Wickens, M. Targeted translational regulation using the PUF protein family scaffold. *Proc. Natl. Acad. Sci.* 108, 15870–15875 (2011).
77. Sprengel, R. & Hasan, M. T. Tetracycline-controlled genetic switches. *Handb. Exp. Pharmacol.* 178, 49–72 (2007).
78. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.* 7, 215–9 (2000).
79. Ramos, J. L. *et al.* The TetR Family of Transcriptional Repressors The TetR Family of Transcriptional Repressors. *Microbiol. Mol. Biol. Rev.* 69, 326–356 (2005).
80. Tiebel, B., Aung-Hilbrich, L. M., Schnappinger, D. & Hillen, W. Conformational changes necessary for gene regulation by Tet repressor assayed by reversible disulfide bond formation. *EMBO J.* 17, 5112–5119 (1998).
81. Berens, C. & Hillen, W. Gene regulation by tetracyclines: Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *Eur. J. Biochem.* 270, 3109–3121 (2003).
82. Hunsicker, A. *et al.* An RNA Aptamer that Induces Transcription. *Chem. Biol.* 16, 173–180 (2009).
83. Müller, M., Weigand, J. E., Weichenrieder, O. & Suess, B. Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res.* 34, 2607–2617 (2006).
84. Steber, M., Arora, A., Hofmann, J., Brutschy, B. & Suess, B. Mechanistic basis for RNA aptamer-based induction of TetR. *ChemBioChem* 12, 2608–2614 (2011).
85. Tiebel, B. *et al.* Domain motions accompanying Tet repressor induction defined by changes of interspin distances at selectively labeled sites. *J. Mol. Biol.* 290, 229–240 (1999).
86. Goldfless, S. J., Belmont, B. J., De Paz, A. M., Liu, J. F. & Niles, J. C. Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. *Nucleic Acids Res.* 40, (2012).
87. Goldfless, S. J., Wagner, J. C. & Niles, J. C. Versatile control of Plasmodium falciparum gene expression with an inducible protein–RNA interaction. *Nat. Commun.* 5, 5329 (2014).
88. Ausländer, D., Wieland, M., Ausländer, S., Tigges, M. & Fussenegger, M. Rational design of a small molecule-responsive intramer controlling transgene expression in mammalian cells. *Nucleic Acids Res.* 39, (2011).
89. Will, C. L. & Lührmann, R. Spliceosome structure and function. TL - 3. *Cold Spring Harb. Perspect. Biol.* 3 VN-re, 1–23 (2011).
90. Wahl, M. C., Will, C. L. & Lührmann, R. The Spliceosome: Design Principles of

a Dynamic RNP Machine. *Cell* 136, 701–718 (2009).
91. Pérez-Valle, J. & Vilardell, J. Intronic features that determine the selection of the 3′ splice site. *Wiley Interdiscip. Rev. RNA* 3, 707–717 (2012).
92. Faustino, N. A., Cooper, T. a & Andre, N. Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437 (2003).
93. Ast, G. How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773–782 (2004).
94. SM, B. Exon recognition in vertebrate splicing. *Mol Cell Biol* 2411–4 (1995).
95. Smith, C. W. J. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388 (2000).
96. Chabot, B. Directing alternative splicing: Cast and scenarios. *Trends Genet.* 12, 472–478 (1996).
97. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32 (2015).
98. Venables, J. P. Aberrant and Alternative Splicing in Cancer Aberrant and Alternative Splicing in Cancer. 64, 7647–7654 (2004).
99. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.* 579, 1900–1903 (2005).
100. Baralle, D., Lucassen, A. & Buratti, E. Missed threads. *Organization* (2009). doi:10.1038/embor.2009.170
101. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and Disease. *Cell* 136, 777–793 (2009).
102. Warf, M. B., Diegel, J. V, von Hippel, P. H. & Berglund, J. A. The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 9203–9208 (2009).
103. Groher, F. Kontrolle des prä-mRNA Spleißens durch synthetische Riboswitche.Dostoral thesis. 2015. TU Darmstadt, Department of Biology.
104. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* 6, 451–464 (2005).
105. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: From microbes to mammals. *Cell* 144, 910–925 (2011).
106. Roßmanith, J. & Narberhaus, F. Modular arrangement of regulatory RNA elements. *RNA Biol.* 14, 287–292 (2017).
107. Rimoldi, V. *et al.* Dual Role of G-runs and hnRNP F in the Regulation of a Mutation-Activated Pseudoexon in the Fibrinogen Gamma-Chain Transcript. *PLoS One* 8, 1–11 (2013).
108. Han, J. *et al.* SR Proteins Induce Alternative Exon Skipping through Their Activities on the Flanking Constitutive Exons. *Mol. Cell. Biol.* 31, 793–802 (2011).
109. Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci. Rep.* 5, 16273 (2015).
110. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* 7, 2193 (2017).
111. Morris, D. R. & Geballe, a P. Upstream Open Reading Frames as Regulators of

mRNA Translation. *Mol. Cell. Biol.* 20, 8635–8642 (2000).
112. Sachs, M. S. & Geballe, A. P. Downstream control of upstream open reading frames. *Cold Spring Harb. Lab. Press* 20, 915–921 (2006).
113. Hinnebusch, A. G. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* 75, 434–467 (2011).
114. Hill, R., Cautain, B., de Pedro, N. & Link, W. Targeting nucleocytoplasmic transport in cancer therapy. *Oncotarget* 5, (2014).
115. Bauer, N. C., Doetsch, P. W. & Corbett, A. H. Mechanisms Regulating Protein Localization. *Traffic* 16, 1039–1061 (2015).
116. Hung, M.-C. & Link, W. Protein localization in disease and therapy. *J. Cell Sci.* 124, 3381–3392 (2011).
117. Edalat, F. NIH Public Access. 40, 1301–1315 (2012).
118. Kim, Y. H., Han, M.-E. & Oh, S.-O. The molecular mechanism for nuclear transport and its application. *Anat. Cell Biol.* 50, 77 (2017).
119. Reverdatto, S. *et al.* Subcellular Protein Localisation in Health and Disease. *Nat. Commun.* 5, 1–7 (2016).
120. Abil, Z., Gumy, L. F., Zhao, H. & Hoogenraad, C. C. Inducible Control of mRNA Transport Using Reprogrammable RNA-Binding Proteins. *ACS Synth. Biol.* 6, 950–956 (2017).
121. Belmont, B. J. & Niles, J. C. Inducible Control of Subcellular RNA Localization Using a Synthetic Protein-RNA Aptamer Interaction. *PLoS One* 7, (2012).
122. Klemm, J. D., Beals, C. R. & Crabtree, G. R. Rapid targeting of nuclear proteins to the cytoplasm. *Curr. Biol.* 7, 638–644 (1997).
123. Kudo, N. *et al.* Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc. Natl. Acad. Sci.* 96, 9112–9117 (1999).
124. Kakar, M., Davis, J. R., Kern, S. E. & Lim, C. S. Optimizing the protein switch: Altering nuclear import and export signals, and ligand binding domain. *J. Control. Release* 120, 220–232 (2007).
125. Beyer, H. M. *et al.* Red Light-Regulated Reversible Nuclear Localization of Proteins in Mammalian Cells and Zebrafish. *ACS Synth. Biol.* 4, 951–958 (2015).
126. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* 5, 1–11 (2014).
127. Zillmann, M., Zapp, M. L. & Berget, S. M. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol. Cell. Biol.* 8, 814–21 (1988).
128. Meyer, M., Plass, M., Pérez-Valle, J., Eyras, E. & Vilardell, J. Deciphering 3'ss Selection in the Yeast Genome Reveals an RNA Thermosensor that Mediates Alternative Splicing. *Mol. Cell* 43, 1033–1039 (2011).
129. Vigevani, L., Gohr, A., Webb, T., Irimia, M. & Valcárcel, J. Molecular basis of differential 3′ splice site sensitivity to anti-tumor drugs targeting U2 snRNP. *Nat. Commun.* 8, 1–15 (2017).
130. Wu, W. *et al.* Alternative splicing controls nuclear translocation of the cell cycle-

regulated Nek2 kinase. *J. Biol. Chem.* 282, 26431–26440 (2007).
131. Cassago, A. *et al.* Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. *Proc. Natl. Acad. Sci.* 109, 1092–1097 (2012).
132. Hanson, S., Bauer, G., Fink, B. & Suess, B. Molecular analysis of a synthetic tetracycline-binding riboswitch Molecular analysis of a synthetic tetracycline-binding riboswitch. *RNA* 11, 503–511 (2005).
133. Koloteva, N., Müller, P. P. & McCarthy, J. E. G. The position dependence of translational regulation via RNA-RNA and RNA- protein interactions in the 5'-untranslated region of eukaryotic mRNA is a function of the thermodynamic competence of 40 S ribosomes in translational initiation. *J. Biol. Chem.* 272, 16531–16539 (1997).
134. Suess, B. *et al.* Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res.* 31, 1853–1858 (2003).
135. Goldfless, S. J., Belmont, B. J., De Paz, A. M., Liu, J. F. & Niles, J. C. Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. *Nucleic Acids Res.* 40, 1–12 (2012).
136. Groher, F. *et al.* Riboswitching with ciprofloxacin—development and characterization of a novel RNA regulator. *Nucleic Acids Res.* 46, 2121–2132 (2018).
137. Davuluri, R. V, Suzuki, Y., Sugano, S. & Zhang, M. Q. CART Classification of Human 5 ′ UTR Sequences CART Classification of Human 5J UTR Sequences. *Genome Res.* 1807–1816 (2000). doi:10.1101/gr.GR-1460R
138. Kötter, P., Weigand, J. E., Meyer, B., Entian, K.-D. & Suess, B. A fast and efficient translational control system for conditional expression of yeast genes. *Nucleic Acids Res.* 37, e120–e120 (2009).
139. Del Vecchio, D., Ninfa, A. J. & Sontag, E. D. Modular cell biology: Retroactivity and insulation. *Mol. Syst. Biol.* 4, (2008).
140. Pantoja-Hernández, L. & Martínez-García, J. C. Retroactivity in the Context of Modularly Structured Biomolecular Systems. *Front. Bioeng. Biotechnol.* 3, 85 (2015).
141. Buratti, E. & Baralle, F. E. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process MINIREVIEW Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol. Cell. Biol.* 24, 10505–10514 (2004).
142. Taggart, A. J., Desimone, A. M., Shih, J. S., Filloux, M. E. & Fairbrother, W. G. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. *Nat. Struct. Mol. Biol.* 19, 719–721 (2012).
143. Horowitz, D. S. The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* 3, 331–350 (2012).
144. Reich, C. I., VanHoy, R. W., Porter, G. L. & Wise, J. A. Mutations at the 3′ splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. *Cell* 69, 1159–1169 (1992).
145. Hallegger, M., Sobala, A. & Smith, C. W. J. Four exons of the serotonin receptor 4 gene are associated with multiple distant branch points. 1–13 (2010).

doi:10.1261/rna.2013110.10

146. Jenkins, J. L., Agrawal, A. A., Gupta, A., Green, M. R. & Kielkopf, C. L. U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res.* 41, 3859–3873 (2013).
147. Martin, R. M. *et al.* Principles of protein targeting to the nucleolus. *Nucleus* 6, 314–325 (2015).
148. Hink, M. A. *et al.* Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. *J. Biol. Chem.* 275, 17556–17560 (2000).
149. Lee, Y. J., Wei, H. M., Chen, L. Y. & Li, C. Localization of SERBP1 in stress granules and nucleoli. *FEBS J.* 281, 352–364 (2014).
150. Karniely, S. & Pines, O. Single translation-dual destination: Mechanisms of dual protein targeting in eukaryotes. *EMBO Rep.* 6, 420–425 (2005).
151. Fukasawa, Y. *et al.* MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol. Cell. Proteomics* 14, 1113–1126 (2015).
152. Koehler, C. M. Protein translocation pathways of the mitochondrion. *FEBS Lett.* 476, 27–31 (2000).
153. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci.* 83, 2850–2854 (1986).
154. Kozak, M. Features in the 5′ non-coding sequences of rabbit α and β-globin mRNAs that affect translational efficiency. *J. Mol. Biol.* 235, 95–110 (1994).
155. Hentze, M. W. *et al.* Identification of the iron responsive element for the translational regulation of human ferritin mRNA. *Science (80-.).* 238, 1570–1573 (1987).
156. Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Curr. Opin. Struct. Biol.* 13, 56–63 (2003).
157. Gebauer, F. & Hentze, M. W. Molecular mechanisms of translational control. *Nat. Rev. Mol. Cell Biol.* 5, 827–835 (2004).
158. Babendure, J. R., Babendure, J. L., Ding, J. & Tsien, R. Y. Control of mammalian translation by mRNA structure near caps Control of mammalian translation by mRNA structure near caps. *Rna* 12, 851–861 (2006).
159. Schneider, C., Bronstein, L., Diemer, J., Koeppl, H. & Suess, B. ROC'n'Ribo: Characterizing a Riboswitching Expression System by Modeling Single-Cell Data. *ACS Synth. Biol.* 6, 1211–1224 (2017).
160. Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. G. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. *Mol. Microbiol.* 49, 1627–1637 (2003).
161. Ganesan, S. M., Falla, A., Goldfless, S. J., Nasamu, A. S. & Niles, J. C. Synthetic RNA–protein modules integrated with native translation mechanisms to control gene expression in malaria parasites. *Nat. Commun.* 7, 10727 (2016).
162. Kopniczky, M. B., Moore, S. J. & Freemont, P. S. Multilevel Regulation and Translational Switches in Synthetic Biology. *IEEE Trans. Biomed. Circuits Syst.* 9, 485–496 (2015).

163. Qi, L. S. *et al.* NIH Public Access. *Curr. Opin. Cell Biol.* 152, 1173–1183 (2013).
164. Agne, M. *et al.* Modularized CRISPR/dCas9 effector toolkit for target-specific gene regulation. *ACS Synth. Biol.* 3, 986–989 (2014).
165. Lohmueller, J. J., Armel, T. Z. & Silver, P. A. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucleic Acids Res.* 40, 5180–5187 (2012).
166. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res.* 40, 7584–7595 (2012).
167. Olson, E. J. & Tabor, J. J. Post-translational tools expand the scope of synthetic biology. *Curr. Opin. Chem. Biol.* 16, 300–306 (2012).
168. Westbrook, A. M. & Lucks, J. B. Achieving large dynamic range control of gene expression with a compact RNA transcription-translation regulator. *Nucleic Acids Res.* 45, 5614–5624 (2017).
169. Re, A. Synthetic Gene Expression Circuits for Designing Precision Tools in Oncology. *Front. Cell Dev. Biol.* 5, 1–9 (2017).
170. Engstrom, M. D. & Pfleger, B. F. Transcription control engineering and applications in synthetic biology. *Synth. Syst. Biotechnol.* 2, 176–191 (2017).
171. Ip, J. Y. *et al.* Global analysis of alternative splicing during T-cell activation. *Spring* 563–572 (2007). doi:10.1261/rna.457207.attenuated
172. Arpino, J. A. J. *et al.* Tuning the dials of synthetic biology. *Microbiol. (United Kingdom)* 159, 1236–1253 (2013).
173. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, 45e–45 (2001).
174. Berens, C. *et al.* Subtype selective tetracycline agonists and their application for a two-stage regulatory system. *ChemBioChem* 7, 1320–1324 (2006).

9 Talks and Poster presentations

2 nd International Caparica Conference in Splicing, Caparica-Lisbon, Portugal **2018**
Oral and Poster presentation

MetaRNA: *Discussions about science and the future*, Darmstadt, Germany **2017**
Oral presentation

The 22nd Annual Meeting of the RNA Society, Prague, Czech Republic **2017**
Poster presentation

Aptamers 2017: 4rd Oxford symposium on Aptamers, Oxford, UK **2017**
Poster presentation

MetaRNA: *Midterm review*, Copenhagen, Denmark **2017**
Oral presentation

MetaRNA: *Metabolism at the single-cell level*, Groningen, Netherlands **2016**
Oral presentation

1 st International Caparica Conference in Splicing, Caparica-Lisbon, Portugal **2016**
Poster presentation

Aptamers 2016: 3rd Oxford symposium on Aptamers, Oxford, UK **2016**
Poster presentation

MetaRNA: *Riboswitch design*, Darmstadt, Germany **2016**
Oral presentation

MetaRNA: *Aptamer selection and characterization*, Bordeaux, France **2015**
Oral presentation

FEBS-EMBO 2014 Conference, Paris, France **2014**
Poster presentation

14th FEBS Young Scientists' Forum, Paris, France **2014**
Poster presentation

XXI Molecular Biology Meeting, Barcelona, Spain **2014**
Oral presentation

Barcelona BioMed Conference, Barcelona, Spain **2013**
Poster presentation

3rd IRB PhD Student Symposium, Barcelona, Spain **2013**
Poster presentation

10 Publications

Mol, A. A., M. Vogel and B. Suess. 2020. Inducible nuclear import by TetR aptamer-controlled 3' splice site selection. RNA DOI: 10.1261/rna.077453.120

Mol, A. A., F. Groher, B. Schreiber, C. Rühmkorff and B. Suess. 2019. Robust gene expression control in human cells with a novel universal TetR aptamer splicing module. Nucleic Acids Research, 47: https://doi.org/10.1093/nar/gkz753

Mol, A. A. and M. Stolarek. 2011. Induced damages of DNA by mutagens and the mechanisms of their formation. Advances in Cell Biology 38: 491-505

11 *Curriculum Vitae*

Name: Adam Artur Mol
Date of birth: 24th December 1988 in Tarnowskie Góry, Poland
Address: Waldstrasse 43, 64354 Reinheim, Germany

PROFESSIONAL EXPERIENCE

Doctoral Candidate **2015-2018**
Research group: Synthetic Genetic Circuits, Department of Biology
Technical University of Darmstadt (TUD), Darmstadt, Germany
Doctoral thesis: A TetR-binding aptamer as versatile regulatory element
Supervisor: Prof. Beatrix Suess

Graduate Student **2012-2014**
Research group: Molecular mechanisms of pre-mRNA splicing
Department of Molecular Genomics,
Molecular Biology Institute of Barcelona (IBMB-CSIC), Barcelona, Spain
Supervisor: Dr. Josep Vilardell

Intern **2012**
State Sanitary-Epidemiological Inspection Bytom, Poland
Organization and performance of sanitary and epidemiological examinations, surveys and tests, toxicological, and hygienic assessments and other types of expert analysis.

Erasmus Intern **2011**
Research group: Control, evolution and genetic impact of plant transposons
Department of Molecular Genetics
The Center for Research in Agricultural Genomics (CRAG), Barcelona, Spain
Supervisor: Dr. Josep Casacuberta

Trainee **2009**
Department of Nuclear Medicine and Endocrine Oncology
Centre of Oncology Maria Skłodowska-Curie Memorial Institute, Gliwice, Poland
Participation in research focused on endocrine-related cancer, especially thyroid, parathyroid, adrenal glands and pituitary
Supervisor: Prof. M.D. Barbara Jarząb

EDUCATION

M.Sc. in Biotechnology **2010-2012**
Department of Genetics, University of Silesia (US), Katowice, Poland
Master thesis: *In silico* analysis and cloning of *Hordeum vulgare AtNCED3*, *AtAHG2* and *AtSAD1* homologs
Supervisor: Prof. Dr. Mirosław Małuszyński and Dr. Agata Daszkowska-Golec
Graduated Summa Cum Laude with perfect grade (6)

Postgraduate Pedagogical Course - Biology Teacher **2009-2012**
University of Silesia, Katowice, Poland
Faculty of Biology and Environmental Protection

B.Sc. in Biotechnology **2007-2010**
Department of Genetics, University of Silesia (US), Katowice, Poland
Bachelor thesis: Mutagen-induced damages of DNA and the mechanisms of their formation
Supervisor: Prof. Dr. Mirosław Małuszyński

12 Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 31.05.2018

.....
(Adam Artur Mol)